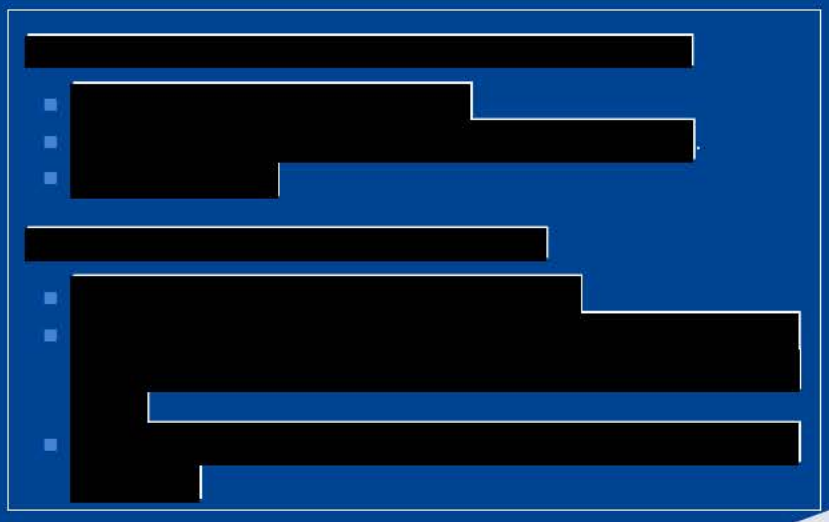
14/12/2022

AI in CRRS in the Context of ETIAS and Revised VIS Final Report



The report does not reflect eulisa position and future decision on the topic. It was commission to support the reflection of the agency on the topic.

Table of Contents

List of Tables

Table of Figures

Executive Summary

Organisations today are required to combine ever-more complex strategies to take short window of time business decisions based on vast amounts of data. In this regard, Artificial Intelligence (AI) can be used to support decision-making and results accuracy via identifying patterns, drawing conclusions, making case-by-case choices and others. Although Artificial Intelligence research has been ongoing for decades, it can be considered a relatively new mainstreamed discipline or concept; as the amount of data available has continuously grown, AI models are continuously being developed/refined, tested and put in production and new AI environments, tools and techniques, are being designed and made available.

The projects under the eu-LISA umbrella are primarily concerned with homeland internal security, and their analysis needs can be met by utilizing a deterministic approach (the process of calculating a future event accurately, without involving randomness). For those analytical needs where humans take decisions, opportunities exist to enhance the current process and assist the decision maker / system user, by applying Artificial Intelligence tools and technics to:

- Identify significant patterns within the data.
- Provide statistical tools to analyse data.
- Support users with the analysis of current and historical data.
- Provide assistance with repetitive tasks.

This high-level Study showcases the AI potential and the readily accessible AI tools and technology that can be applied today; it is not, however, a comprehensive detailed analysis and design document that typically precedes implementation. The detailed analysis and design are needed, as the next step, prior to any deployment of AI tools and solutions to eu-LISA. Furthermore, it is important to emphasize this is a research document and the outcome does not indicate that the AI Solutions will be utilized in the manner suggested and described in the Study

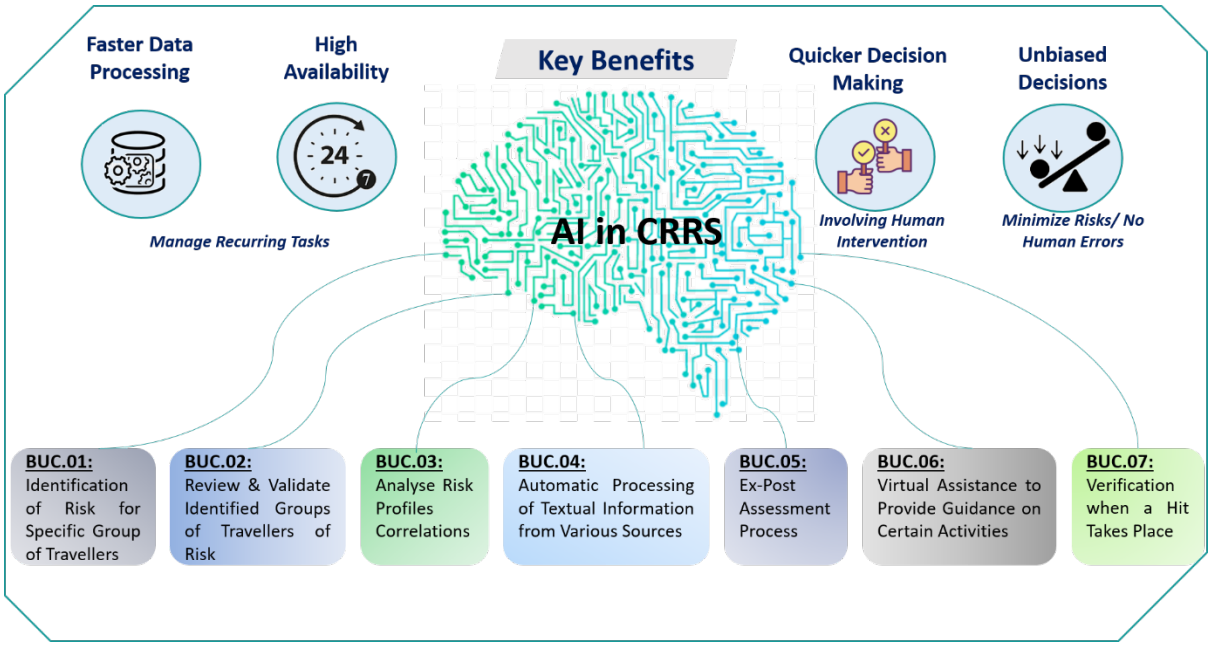
The Study Objective

This Study aspires to highlight how Artificial Intelligence Solutions (AI, tools and techniques) can be utilised to support CRRS, ETIAS and Revised VIS. It proposes AI Solution enhancements to the current processes within risk profiling, screening rules and risk indicators, which require a high level of human intervention involving decision-making. The Study examines and identifies seven Business Use Cases and suggests a plethora of Open-Source AI tools and frameworks, with the most potential for the AI use to support CRRS, ETIAS and Revised VIS.

The Study takes into account the broader perspective of the current eu-LISA IT landscape, and proposes the design of AI-enabled frameworks, i.e., Data Science and Machine Learning (DSML) and platforms (as described on Chapter 5 – *Applicable AI Technologies & Assessment in CRRS*). The recommended AI-enabled frameworks provide end-to-end data management lifecycle and related processes, with a focus on how data will be stored, processed, and finally used by the authorities.

Additional effort and focus are given to how the AI-enabled model outcomes will be communicated, to the authorised users / authorities, rather than to simply expose AI and Machine Learning (ML) methodologies.

Business Use Cases & Proposed AI solutions The Business Use Cases identified as candidates for implementation of AI solutions are grouped as per the respective IT system: CRRS and ETIAS.



The above figure illustrates the key benefits of AI for each Business Use Case in the context of CRRS and ETIAS

- **CRRS (Central Repository for Reporting and Statistics):** collects and stores statistical data from VIS/ EES/ ETIAS of over-stayers and information regarding refusal of entry of Third Country National applicants' information. Business Use Cases identified for this reporting system and the respective findings / AI recommendations are:

    - **BUC 1**: Identification of risk for a specific group of travellers.

The AI technology will enhance this process by identifying patterns or a set of common characteristics from the analysis of historical data in the CRRS (collected from the Core Business Systems -CBS) related to security, illegal immigration or high epidemic risk for a specific group of travellers (Clusters).

    - **BUC 4**: Automatic processing of textual information from various sources.

The AI technology will enhance the current process by detecting common textual characteristics in manually uploaded data and will facilitate the identification of specific groups of travellers of risk (referred to as Clusters).

- **ETIAS (European Travel Information and Authorisation System):** a new IT system, currently under development, which will manage visa travel applications. Business Use Cases identified for this system and the respective findings / AI recommendations are:

    - **BUC 2**: Review and Validate Identified Risk Profile (Ex-Ante).

The AI technology aims to enhance the ex-ante assessment process of screening rules assisting officers in the examination of the application by using stored and historical data and generating an indicator for normal and suspicious outcomes for a specific set of travellers.

    - **BUC 3**: Analyse Risk Profile Correlations.

The AI technology will identify and highlight correlations amongst the risk profiles from its historic data during the ex-ante or ex-post assessment allowing a more precise definition.

    - **BUC 5**: Risk Profile Validity Assessment.

The AI technology will analyse, detect deviations and propose further review of the risk indicators during the ex-post assessment process.

- **BUC 7**: Verification Process When a Hit Takes Place.

The AI technology will support the process of verification when a hit takes place, using ETIAS historical data by implementing model training of the corresponding risk profiles.

- **Virtual Assistance:** An IT area that eu-LISA is considering for future implementation.
  - **BUC 6**: Chatbot / Virtual Assistance.

The AI technology will reply to contextual support requests and offer assistance for crucial business processes; some examples of which are mentioned in the Business Use Case.

The Study Conclusion

Based on the analysis of business use cases, the use of AI frameworks can significantly improve the business processes, under review, and improve timely decision-making by users. Several of the current AI tools and techniques, available in the field of Open-Source[1] solutions, are sufficiently mature and can be used relatively easily and safely and put in production at eu-LISA; the AI Solutions proposed in the study can be found on ***Appendix I - Proposed Technological Frameworks.***

In terms of human resources, roles and activities, a sizing exercise has been proposed, based on the identified AI Frameworks for the deployment of the complete list of Business Use Cases; while indicative estimations are reported on ***Appendix II – Assessment of the Foreseen Implementation Effort*** The respective sizing exercise articulates an estimate of 2050 man-days, under the "most probable" scenario presented. The effort may be organized to include approximately 24 Sprints or incremental deliveries (a short, defined period, usually *under a 2-week effort used to organize work cycles*), running either sequentially or in parallel, depending always on the availability of specialists / resources, interfaces to other systems, and further organizational/systemic interdependencies, *as outlined in sub-section 8.4.3 Indicative Time Plan of the final Report Document*.

Innovative AI frameworks can support business processes that involve decision making by evaluating and analysing data faster, when compared to traditional data processing tools and technologies. AI-enabled systems typically support high availability in managing recurring data processing tasks, on a 24/7 basis. Apart from this, another key benefit of AI Solutions, lies in their capability to adapt to changes, in order to meet the organisation's future demands.

The next stage as a follow-up to this study, could be to conduct an AI Frameworks Detailed Analysis and Design project to refine the Business Use Case requirements and to develop the IT infrastructure as well as the interfaces necessary with the related systems.

The Methodology

The methodological approach applied in this Study is **DMADV**[2] (Define - Measure - Analyse - Design - Verify); a Six Sigma framework that provides organisations with the necessary tools to improve their business processes. Additionally, the Study endeavours to provide a high-level project duration and implementation effort estimation. The technique used for duration and implementation estimation is the Program Evaluation and Review Technique

[1] Further information of Open-Source can be found at https //opensource dev/
[2] Further information on DMADV and Six Sigma can be found at https //www sixsigmadaily com/what-is-dmadv/

(**PERT** method). PERT[3] provides optimistic, average and pessimistic effort estimation models, based on the key assumptions of resource availability, potential dependencies and organisational capacity. For more information regarding the methodologies and effort assessment, the readers may refer to ***Chapter 5 – Applicable AI Technologies & Assessment in CRRS***, and *Appendix II – Assessment of the Foreseen Implementation Effort*

Accountability and Legal Framework

In 2020 eu-LISA produced the report: "Artificial Intelligence in the Operational Management of Large-scale IT Systems" which explores the possibilities of using artificial intelligence technologies in the context of eu-LISA business processes. One of the areas where AI Solutions can be applied, in the context of the report, is ETIAS (European Travel Information and Authorisation System) a form of electronic travel authorisation or visa waiver. Under this framework, the identification of risk profiles was crucial in this Study, given its direct impact but isolated from other areas, and many of the identified Business Use Cases orbit around this.

Directly involved Stakeholders for this Study are: Frontex, FRA, and European Commission, eu-LISA (CRRS, ETIAS and Recast VIS). The Study followed a 3 phased incremental approach. Once Business Use Cases were defined and agreed, the Study conducts a comprehensive evaluation of the AI solutions that would enable the Business Use Cases, whilst adhering to business and stakeholder requirements.

The Study is designed in accordance with the Accountability Principles Framework (AP4AI) Blueprint which evaluates and enforces legal and appropriate use of AI. These principles are:

- **Legality**: all aspects of the use of AI should be lawful and be governed by formal, promulgated rules. It extends to all those involved in building, developing and operating AI systems for use in a criminal justice context.
- **Universality**: all relevant aspects of AI deployments within the internal security community are covered through the accountability process. This includes all processes: design, development and supply, domains, aspects of police mission, AI systems, and stages in the AI lifecycle or usage purposes.
- **Pluralism**: ensures that oversight involves all relevant stakeholders engaged in and affected by a specific AI deployment. Pluralism avoids homogeneity and thus a tendency or perception for the regulators to take a one-sided approach.
- **Transparency**: involves making available clear, accurate and meaningful information about AI processes and specific deployment pertinent for assessing and enforcing accountability.
- **Independence**: it refers to the status of competent authorities performing oversight functions in respect of achieving accountability.
- **Commitment to Robust Evidence**: documented records or other proof of compliance measures in respect of legal and other formal obligations pertaining to the use of AI in an internal security context.
- **Enforceability and Redress**: requires mechanisms to be established that facilitate independent and effective oversight in respect of the use of AI in the internal security community, as well as mechanisms to respond appropriately to instances of non-compliance with applicable obligations by those deploying AI in a criminal justice context.
- **Compellability**: the need for competent authorities and oversight bodies to compel those deploying or

[3] Further information on PERT can be found at https //www investopedia com/terms/p/pert-chart asp

utilising AI in the internal security community to provide access to necessary information, systems or individuals by creating formal obligations in this regard.

- **Explainability**: requires those using AI to ensure that information about this use is provided in a meaningful way that is accessible and easily understood by the relevant participants/audiences.
- **Constructiveness**: embraces the idea of participating in a constructive dialogue with relevant stakeholders involved in the use of AI and other interested parties, by engaging with and responding positively to various inputs.
- **Conduct**: governs how individuals and organisations will conduct themselves in undertaking their respective tasks and relates to sector-specific principles, professional standards and expected behaviours relating to conduct within a role, which incorporate integrity and ethical considerations
- **Learning Organisation**: promotes the willingness and ability of organisations and people to improve AI through the application of (new) knowledge and insights.

Structure of the Study

The Study is organised in 6 chapters with 2 appendixes providing additional supporting information:

- Chapters 1 and 2: present the reader with an introduction, as well as the methodology used to conduct the Study.
- Chapters 3 and 4: focus on the actual Business Use Cases selected, as candidates to AI solution improvements, and outline the respective business and stakeholder requirements.
- Chapter 5: examines and provides an insight on the various AI Solutions (tools and technologies) available and proposed from the Open-Source world.
- Chapter 6: presents conclusions and recommendations with respect to the Study findings.
- Appendix 1: provides information on the proposed technological framework in consideration with the existing eu-LISA technical environment.
- Appendix 2: provides a high-level estimation of the foreseen implementation effort using an industry standard methodology (PERT).

Brief Definition of Terms

Term	Definition
Risk Profile	**Set of specific risk indicators associated with a certain weight that can produce a risk condition outcome related to security or illegal immigration or a high epidemic risk**.
Screening Rules	The ETIAS screening rules shall be an algorithm enabling profiling as defined in point 4 of Article 4 of Regulation (EU) 2016/679 through the comparison in accordance with Article 20 of this Regulation of the data recorded in an application file of the ETIAS Central System with specific risk indicators established by the ETIAS Central Unit under paragraph 4 of this Article pointing to security, illegal immigration or high epidemic risks.
Risk Indicators	A combination of data established by the ETIAS Central Unit that might drive to a risk condition related to security or illegal immigration or a high epidemic risk, including one or several of the following: (a) Age range, sex, nationality;

(b) country and city of residence;
(c) level of education (primary, secondary, higher or none);
(d) current occupation (job group).

1 Chapter 1 - Introduction

1.1 Objective and Scope

The objectives of this study are formulated as follows:

1. Identify on the business use cases and elicit business and stakeholder requirements for the use of Artificial Intelligence (AI) in CRRS in the context of ETIAS and Revised VIS, specifically to support the risk profiling/ screening rules/ risk indicators.
2. Perform a thorough assessment of the AI solutions that enable the defined Business Use Cases adhering to the business and stakeholders' requirements.
3. Produce a high-level assessment of the foreseen implementation effort of the proposed AI solutions.

The current study has been delivered within three phases:

- Phase 1, corresponds to the first objective of the study, i.e. the identification of Business Use Cases (BUCs) for the use of AI in CRSS in the context of ETIAS and Revised VIS, specifically to support the risk profiling/ screening rules/ risk indicators;
- Phase 2, corresponds on producing the detailed business and stakeholders' requirements;
- Phase 3 involves recommending the most suitable AI solutions that allow the Business Use Cases to deliver results that can be taken action on, in accordance with the requirements.
- Elicitation of Functional and Non-Functional requirements.
- Assessment on the significance of the impact of AI in the current capabilities in CRRS/ETIAS.

1.2 Document Description

This document is divided into five main Chapters, and two Appendixes. Chapter 1 contains the introductory part of the current study, while it provides an overview of the project's approach, which is divided into three Phases. Additionally, the initial steps of the proposed AI solution are first presented on Chapter 1. The protection of fundamental rights in the content of AI, the overall activities and time-plan, the intended audience, the legal base and related documentation, and finally the definition of terms that are widely used in this study, are also main topics of Chapter 1. Chapter 2 is dedicated to the description of information gathering methodology for determining specific Business Use Cases that are of special interest for eu-LISA and its stakeholders. Additionally, Chapter 2 contains the requirements gathering methodology, and the methodology for identifying the proper AI tools. Chapter 3 contains the complete description of the Business Use Cases, which are eligible for this study. It is crucial to mention that the Business Use Cases (BUCs) candidates were discussed before their final determination, during three workshops: (i) First one with FRONTEX on the 29th of March 2022; (ii) Second one with eu-LISA on the 5th of May 2022; (iii) Third one with eu-LISA, for technical deliberation, on the 13th of May 2022.

Chapter 4 includes the requirements hierarchy applying for the different BUCs; while the respective stakeholders' and business requirements are presented on Appendix III – Business and Stakeholder Requirements. Chapter 5 is dedicated to the applicable AI technologies and their assessment. First, an overview of CRRS architectural landscape is implemented, and then the methodology for identifying the open-source AI tools is refined in order to be applicable for the specifications of Phase 3. Additionally, this study proposes an innovative Data Science and Machine Learning platform that can host and enable via novel AI solutions, the various BUCs. Finally, the

Consortium provides an in-depth analysis and assessment of the various BUCs, by suggesting the best-fit solution for each one of them. Chapter 6 illustrates the conclusions of the current study, and further provides Consortium's recommendations to eu-LISA. Chapter 7 includes all technological frameworks that are considered for the current study, while Chapter 8 (Appendix II – Assessment of the Foreseen Implementation Effort) provides the assessment of the foreseen implementation effort for this project, an indicative project plan with certain work packages, and finally an indicative project team, along with their allocations and effort scheme.

1.3 Approach

This study is delivered in three phases. Each one of them includes specific deliverables and activities described below:

1.3.1 Phase 1 - Inception Report - Activities and Deliverables:

Conducted Activities:

- Kick-off meeting planning and execution.
- Initial analysis and assessment of relevant documentation.
- Proposal presented to eu-LISA with the approach to conduct the study, including but not limited to workshops to gather Business Use Cases (BUCs) and requirements and establish interdependence between BUC and AI tools, etc.
- Agreement with eu-LISA on the approach and specific timelines to perform the study.
- Agreement with eu-LISA on the list of BUC's analysed more extensively during Phase 2.
- Preparation of Inception Report, which goes through the formal deliverable review process as defined in the FQP (Framework Quality Plan).

Deliverable: "Inception Report"

The delivery for this phase is the Inception Report, this document includes an introduction, the proposed study approach to eu-LISA and an appendix with the list of Business Use Cases (BUCs) elaborated in following phases. Therefore:

- Chapter 1 – Introduction.
- Chapter 2 – Methodology.
- Chapter 3 – Proposed Business Use Cases.
- Appendix.

1.3.2 Phase 2 – Interim Report – Activities and Deliverables:

Conducted Activities:

- Analysis of the relevant documentation.
- Analysis of the Frontex' processes concerning screening rules.

The anticipated line of work is performed through workshops with the involved stakeholders covering the following aspects:

- Requirements gathering.
- Technical workshops and meetings.
- Requirements elicitation per the BABOK guide (Business Analysis Body of Knowledge).

Activities specific to AI aspects and based on DMADV methodology (Define, Measure, Analysis, Design, and Verify)

are performed, including steps such as:

- Problem definition.
- Measure.

Deliverable: "Interim Report"

The Interim Report, is a deliverable of Phase 2, and includes:

- Chapter 3 - Use cases stemming from legislation and stakeholder consultation.
- Chapter 4 - Business and Stakeholder Requirements, Requirements Traceability Matrix and a separate SPARX EA file (if necessary).
- Chapter 5 – Applicable AI Technologies and Assessment.
- Chapter 6 – Conclusions and Recommendations.

1.3.3 Phase 3 – Final Report – Activities and Deliverables:

Activities include:

Based on agreed Business Use Cases and the elicited requirements explicit to AI on Phase 2, Lot1 conducts a study to identify potential AI technologies and tools to support the Business Use Cases (BUCs) and requirements and follows the steps described in the DMADV methodology (Define, Measure, Analysis, Design, Verify) (as described in **Section 10.5** of this document). The study also includes an assessment of foreseen costs and the required man-days implementation and configuration effort of the proposed solutions and tools.

Deliverable: "Final Report" (This document)

The Final Report is a deliverable of Phase 3 and includes:

- Chapter 4 - Requirements Traceability Matrix and a separate SPARX EA file (if necessary).
- Chapter 5 - A detailed overview together with the outcome of the study for the AI tools/technologies (Applicable/relevant AI tools/technologies/impact assessment).
    - Appendix I – Proposed Technological Frameworks.
    - Appendix II – Assessment of the Foreseen Implementation Effort.
    - Appendix III– Business and Stakeholder Requirements.

Chapter 6 - The conclusions and recommendations, highlighting the essential features and significant outcomes, of all phases (Conclusions & Recommendation).

1.4 The Artificial Intelligence (AI) main drivers

Nowadays, significant changes in technology, industries, and societal patterns have been reported as the main consequence of smart automation and enhanced interconnectivity. On this direction, Artificial Intelligence (AI) has been widely acknowledged as the main driver towards the innovation of several critical technological fields (i.e., including e-Government, transportation, healthcare, manufacturing, commerce, and banking, among others). Moreover, Data Science, Machine Learning, AI and other data-related fields create an excellent opportunity for organisations to optimize their operations, leveraging their massive amount of data to further automate their business processes. According to Gartner, the business value of AI is expected to grow significantly within the next

few years, while 80% of executives believe that automation can be applied to any decision-making process[4],[5]. Consequently, AI has the potential to completely lead the way of progress, to allow organisations to remain competitive and to further pursue new directions with their products and services.

The fundamental criteria of AI that should be considered, when high-quality applications are designed in all technology sectors, are: (i) to incorporate automation of repeatable activities whilst enabling human intervention when necessary, (ii) to extract insights and usable knowledge from data (i.e., intelligence), and (iii) to provide self-awareness in the sense of self-monitoring, analysis and reporting (i.e., smart computing). On this direction, the use of modern and smart technologies enables making smarter and faster decisions regarding the business process, ultimately increasing the productivity and profitability of the overall operation, where AI is recognized as a leading technology in the specific areas.

It is widely acknowledged that most AI systems share several common elements, however they also present significant differences, including the levels of autonomy and sophistication. A common classification of AI is implemented based on the intelligence and sophistication of its features, as follows:

- **Artificial Narrow Intelligence (ANI):** refers to AI systems that can perform specific tasks autonomously with human-like capabilities when they are programmed to do so. Despite ANI systems are widely utilised in several developments, they present a relatively narrow range of competencies.
- **Artificial General Intelligence (AGI):** corresponds to AI systems which are able to replicate entirely human intelligence, including the ability to learn, perceive, understand and function as a human. AGI systems provide the capability of a system to perform fully autonomous and multi-functional learning.
- **Artificial Superintelligence (ASI):** relies on AGI systems' capabilities by exceeding human comprehension. In the case of ASI, further advances in AI arise from super-intelligent AI design improvements to the current intelligent systems.

Unlike the scenarios in Sci-Fi movies, General AI (AGI) has a long future ahead to highlight just part of its capabilities. However, the rumours about AI and how it will allow machines to overrule humans or in a more realistic aspect displace many people from their jobs, are dispersing. Both arguments are alienating the human cognition from the actual contribution of AI. AI aims to release time and effort from humans and provide useful insights by leveraging its computer power. The truth is that AI shall exceed human performance, but only in limited specific tasks that require exhausting computational effort, in which usually vast amount of data are processed, and speed (i.e., in text recognition or image recognition patterns). In case of emotional-related tasks, it is difficult for AI to achieve efficient and trustworthy results that can be compared with the corresponding human ones without guided human intervention[6, 7]

Additionally, AI systems may be categorised with respect to their learning ability, and with the associated techniques that exploit them, for instance Natural Language Processing (NLP), and chatbots, among others, providing different capabilities to the end-users. Specifically, (**Table 1**) provides a short summary of the key technology areas and applications of AI.

[4] Tech Providers 2025 AI Is Transforming Investor Positioning for Tech CEOs Published 3 March 2021 - ID G00735589, By Analyst(s) Patrick Stakenas, Evan Brown, David Adams, Alastair Woolcock, Peter Havart-Simkin, Initiatives Corporate Development for Tech CEOs; https //www gartner com/en/documents/3998853

[5] https //www gartner com/en/information-technology/insights/artificial-intelligence

[6] Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30(1), 99-120.

[7] https //www gartner com/en/information-technology/insights/artificial-intelligence

Category	Technology Area	Description
Methods	Artificial Neural Networks	Artificial neural networks (ANNs) are biologically inspired computational networks that simulate the electrical activity of the human brain and nervous system. These networks emulate a biological neural network, but they use a reduced set of concepts from biological neural systems. Some of the background work for the field of artificial neural networks (ANNs) occurred in the late nineteenth and early twentieth centuries. Deep Neural Networks is a special category of ANNs, by merging several layers of Shallow Neural Networks.
	Machine Learning	Machine Learning (ML) is widely acknowledged as the sub-set of AI which provides the ability to systems to learn independently through the exposure to training data. It stands as the baseline in multiple AI systems. ML can be either: (I) supervised, including the scenario where labelled data is available; (ii) semi-supervised, referring to the case where partially labelled data exist in our datasets; or (iii) unsupervised, where no-labelled information is presented in the training datasets.
	Deep Learning	Deep Learning (DL) allows computational models that are composed of multiple processing layers to learn data representations with multiple levels of abstraction. Specifically, DL discovers complex structures in large data sets by using backpropagation algorithms which indicate how the model should change its internal parameters that are used to compute the representation of the current layer, using the representation from the previous layer.
	Symbolic AI	Contradictory to ML, which focuses on the development of algorithms to enable autonomous problem-solving capabilities, symbolic AI corresponds to the 'explicit embedding of human knowledge and behaviour rules into computer programs. Symbolic systems employ axiomatic inferencing mechanisms and handle discrete symbols, such as rule-based or logic-based systems.
Applications	Natural Language Processing (NLP) & understanding	Natural Language Processing (NLP) encompasses AI related to human language tasks. Specifically, it includes applications that automatically process large amounts of text information, speech-to-text and text-to-speech (i.e., language) generation. It can be also characterized as the automatic, or semi-automatic processing of human language.
	Chatbots/ Conversational Agents	Out of the most rising AI trends in the private and public sector concerns the utilization of applications that interact with users in a conversational format and mimic human conversation, known as "conversational agents", "chatbots" or "bots". Chatbots are intelligent agents, i.e., devices that perceive their environment and take actions that maximize their chance of success for achieving a certain goal. They

Category	Technology Area	Description
		can understand a spoken language and use speech communication as user interface.
	Automated Decision Making	Artificial Intelligence is being used in automated decision making for tasks such as assisting the decision maker to choose certain actions in real-time; to solve stressful decision problems; to reduce the human information overload; to enable up-to-date information; to provide a dynamic response with intelligent agents; and to enable communication required for collaborative decisions, among others.

Table 1. Summary of key technology areas and applications of AI

More specifically, in the context of this study, the adoption of AI solutions is an approach that may enhance the domains of supporting human decisions in defining the screening rules, risk profiles and risk indicators. Via this study, Lot1 proposes a high-level innovative suite of AI solutions, with respect to EU fundamental rights, notably, the right of non-discrimination.

The first step while conducting this study is to consider the needs and comprehend the technological landscape regarding activities such as risk profiling, risk indicators identification and screening rules application. Utilizing the power of the proposed AI solutions, these activities will become further automated and more efficient by leveraging data-driven approaches.

We understand that under the current organization's technology and IT systems environment, a number of Proof of Concepts (POC) have been performed as per feedback; therefore, this study depicts the necessary means to enable AI solutions and tools.

Thus, the main purpose of the study is to analyse state-of-the-art AI solutions and technologies. The anticipated AI algorithms are described in detail in **Chapter 5 – Applicable AI Technologies & Assessment** and are further assessed to select the ones that provide the most appropriate solutions to the BUs that are formed by Lot1 during Chapter (1) and Chapter(2).

1.5 Protection of Fundamental Rights in AI

The use of Artificial Intelligence (AI) in public administration and governance requires strong safeguards. The protection of fundamental rights is paramount to empower citizens and improve governance. The European Union's Charter of Fundamental Rights protects fundamental rights and freedoms across the entirety of the EU, including civil, political, economic and social rights. The Charter was declared in 2000 with the Treaty of Nice and came into force in December 2009 along with the Treaty of Lisbon. The Charter's mission is to reaffirm and promote human rights throughout the EU's borders. Many of the rights contained in the Charter are set out in the European Convention of Human Rights, which protects the human rights of people in countries that belong to the Council of Europe.

Through this study, eu-LISA and the Consortium, intend to offer a reliable solution taking seriously into account fundamental rights at risk, for example the right to human dignity and integrity, the respect of one's private life and the protection of personal data, as well as the right to non-discrimination (in particular against third-country nationals on the grounds of sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual

orientation[8]). There should be extra attention to the vulnerable social groups, especially children. In the suggested solution, one of the main concerns was to respect these values and leverage them towards the envisioned AI landscape of eu-LISA.

By way of clarification, in the context of the ETIAS assessment rules age and sex, mentioned in Article 33 (4)(a) of the ETIAS Regulation 2018/1240 on specific risk indicators, are protected grounds of discrimination (the full list of protected grounds is listed in Article 21 of the EU Charter of Fundamental Rights). Therefore, specific risk indicators should not be based solely, or predominantly, on these grounds. The specific risk indicators mentioned in Article 33 paragraph 4, subparagraphs b to d, are not explicitly covered by the protected grounds of Article 21 of the Charter. However, its application could lead to indirect discrimination by revealing protected grounds of discrimination, for example, occupation may reveal social group.

Finally, highlight the importance of ensuring that the risk indicators and therefore the risk profiles are targeted and proportionate to contribute and ensure the reduction of risks to fundamental rights.

With the purpose of levelling readers in regulatory concepts, it is worth mentioning that this study contemplates the principle of equal treatment meaning that there shall be no direct or indirect discrimination[9]:

(a) direct discrimination shall be taken to occur where one person is treated less favourably than another is, has been or would be treated in a comparable situation on grounds of racial or ethnic origin;

(b) Indirect discrimination shall be taken to occur where an apparently neutral provision, criterion or practice would put persons of a racial or ethnic origin at a particular disadvantage compared with other persons, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary. Hence, the commission has initiated a series of actions including additional funding in the field of innovation and research in AI, applying an AI strategy in EU and an expert-group on AI that will be dedicated to implement the recommendations on AI-related issues on multiple domains.

MS and EU agencies are under obligation to act in line with Fundamental Rights when implementing EU law. Before deploying any machine learning models or applying any analytics processes, one of the most common risk areas is the right to protection of personal data. One of the most important regulations regarding data protection is the General Data Protection Regulation (GDPR). GDPR is one of the strongest privacy and security laws worldwide and it protects the right to privacy, which is part of the 1950 European Convention on Human Rights. Even if it is difficult to combine both the need of data protection with the availability of data for development purposes, by considering data protection as default while developing AI technologies, it is a first step towards the required balance. More specifically, a major aspect in AI that can affect the EU's perception of a respectful and non-discriminate environment is low data quality. Low quality data cannot secure a correct testing process and may lead to an increase of biases and errors in systems. Moreover, poor data quality may have a direct impact on Fundamental Rights especially to the right to non-discrimination since the results of any AI procedure are not trustworthy and cannot be taken into serious consideration.

Coinciding with the publication of this study, the European Union Agency for Fundamental Rights published a report entitled Bias in algorithms - Artificial intelligence and discrimination [10] about the use of artificial intelligence in predictive policing and the detection of offensive speech. The content is related to many areas of this study and is

[8] Article 14 of the Regulation (EU) 2018/1240 of the European Parliament and of the Council of 12 September 2018 establishing a European Travel Information and Authorisation System (ETIAS) - https //eur-lex europa eu/legal-content/EN/TXT/?uri=CELEX%3A02018R1240-20210803 – reinforced by Article 21 of EU Charter of Fundamental Rights

[9] Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin - https //eur-lex europa eu/legal-content/EN/TXT/?uri=celex%3A32000L0043

[10] Bias in algorithms - Artificial intelligence and discrimination | European Union Agency for Fundamental Rights (europa eu)

highly relevant as it goes into detail on key issues such as bias in algorithms (its amplification over time and how it affects people's lives), which can lead to discrimination. Furthermore, it promotes the need for more complete and exhaustive assessments of algorithms in terms of bias before they are used for decision-making that may have an impact on people. Due to time constraints, no changes have been made to this study from the aforementioned FRA report.

1.6 Forward-looking Activities and Approach:

The following plan describes the activities to be performed during phase 1 and phase 2 of the study (latest contract SR16).
Please note that activities and timeline may be revised according to stakeholder availability, eu-LISA inputs and agreement.
The following proposed plan outlines the study execution timeline, phases 1, 2 and the final document review period.

	Inception Phase (SR11)	Interim Report Phase 1 - High Level Timeline			Final Report Phase 2 - High Level Timeline				
		Week 37 12-Sept	Week 38 19-Sept	Week 39 26-Sept	Week 40 03-Oct	Week 41 10-Oct	Week 42 17-Oct	Week 43 24-Oct	Week 44 31-Oct
Phase #0 - Inception Report SfR									
Phase #1 - Report Preparation									
Phase #1 - Interim Report - SfI		16.Sept							
Phase #1 - Interim Report - Acceptance			21.Sept *						
Phase #2 - Report Preparation									
Phase #2 - Final Report submit for SfR				28.Sept *					
Review period									
Authors' position									
Follow-up / Consolidation of comments									
Final Report Closure - ACC									27.Oct. *

1.7 Intended Audience

This document is intended to be used by all stakeholders involved in AI activities for CRRS and ETIAS; and the study development lifecycle, namely (but not limited to):

- eu-LISA research
- eu-LISA Business relations management.
- CRRS representatives (Project Managers, Architects)
- ETIAS representatives (Project Managers, Architects)
- Frontex representatives (Project Managers, Architects)
- FRONTEX
- Europol
- FRA
- Any other authorized stakeholder, who is validated by eu-LISA's and Frontex authorities.

1.8 Legal Base and Documentation

1.8.1 Legal sources

The following legal sources have been used as a framework for the BUCs identification (alongside with the Workshops held and Consultants' expertise) and are used in the following phases as a baseline:

Project	Source	Name	Link	Rationale	Description
ETIAS	02018R1240-20210803	02018R1240-20210803 ETIAS REG	32018R1240	Art. 33	Risk screening framework.
				Art. 84	Use of data for reporting and statistics.
		C(2021)4981-20211123 ETIAS Risks DA	DA Risks 1240	Art. 3(1), Art. 3(2), Art. 3(3),	Analysis of statistical data on overstaying, refusal of entry and refusal of travel authorisation
				Art.4(1)(a)(b)(c), Art.4(2)	Member State analysis of specific security risks or threats
				Art. 5(1)(a)(b)(c), Art.5(2)	Member State analysis of abnormal rates of overstaying and refusals of entry
				Art. 6(2)	Analysis of specific high epidemic risks
				Art. 7	Communication
				Art. 8	Data protection
		X(2022)-20220304 Specification of risks IA	IA risks 1240	Art. 2	Analysis of statistics.
				Art. 3(1), Art. 3(2)	Specification of risks
				Art. 4(1)	Evaluation and review of risks
				Art. 6	Data protection
CRRS	02019R0817-20210803	02019R0817-20210803 Interoperability Visa Borders REG v1	02019R0817	Art. 39(1)	Central repository for reporting and statistics.
			02019R0817	Art. 66(4)	Central repository for reporting and statistics.
	02019R0818-20210803	02019R0818-20210803 Interoperability LEA Immigration REG v1	02019R0818	Art. 39(1)	Central repository for reporting and statistics.
			02019R0818	Art. 62(4)	Reporting and statistics.
	02019R0817-	32021R2223-20210930	32021R2223	Art. 2(5)	Reporting and statistics.

Project	Source	Name	Link	Rationale	Description
	20211215	Operation of CRRS supplementing 2019/817_DA			
	02019R0818-20211215	32021R2222-20210930 Operation of CRRS supplementing 2019/818 DA	32021R2222	Art. 2(5)	Reporting and statistics.
VIS	32021R1134-20210707	VIS Reform Regulation (EU) 2021/1134 32021R1134-20210713 VIS REG	32021R1134	Art.45(a) Art. 9(j)	Use of data for reporting and statistics. Specific risk indicators.
		X(2021)-20211217 Definition of risks DA	DA Risks 767	Art. 3(1), Art. 3(2), Art. 3(3),	Analysis of statistical data on overstaying, refusal of entry and refusal of travel authorisation
				Art.4(1)(a)(b)(c), Art.4(2)	Member State analysis of specific security risks or threats
				Art. 5(1)(a)(b)(c), Art.5(2)	Member State analysis of abnormal rates of overstaying and refusals of entry
				Art. 6(2)	Analysis of specific high epidemic risks
				Art. 7	Communication
				Art. 8	Data protection
		X(2021)-20211217 Specification of risks according to Art. 9j3 IA	IA risks 767	Art. 2	Analysis of statistics
				Art. 3(1), Art. 3(2)	Specification of risks
				Art. 4(1)	Evaluation and review of risks
				Art. 6	Data protection
EES	02017R2226-20210803	02017R2226-20210803 EES REG	02017R2226	Art. 63(1)	Reporting and statistics.
SIS	02018R1862-20210803	02018R1862-20210803 SIS REG	02018R1862	Art. 74(6)	Monitoring and statistics

Table 2. Legal sources table

eu-LISA is part of the **EU Innovation Hub for internal security**, an initiative created by the Council and hosted by Europol, which includes the JHA, JRC and DG HOME agencies. Driven by different initiatives, this study continues

previous proposals such as the project entitled Artificial intelligence (AI) initiatives in ETIAS .

It is important to note that the current versions of the legal texts used do not establish the implementation of solutions based on the use of AI as a requirement.[11]

This study focuses on exploring the existing possibilities always within the current limitations, analysing options and alternatives within the sandbox delimited by the different systems.

1.8.2 Reference Documentation

Other sources of documentation considered:

- [RD01] 02018R1240-20210803 ETIAS REG.
- [RD02] DA Risks - Screening Rules (ETIAS).
- [RD03] IA Risks (Draft) (ETIAS).
- [RD04] 02019R0817-20210803 Interoperability REG.
- [RD05] 02019R0818-20210803 Interoperability REG.
- [RD06] 32021R2223-20210930 Operation of CRRS supplementing 2019/817 DA.
- [RD07] 32021R2222-20210930 Operation of CRRS supplementing 2019/818 DA.
- [RD08] 32021R1134-20210713 VIS REG.
- [RD09] DA Risks - Screening rules (Draft) (VIS).
- [RD10] IA Risks (Draft) (VIS).
- [RD11] ETIAS Risk Screening BRD v.1.1 (2).
- [RD12] ETIAS Risk Screening UX v1.1.
- [RD13] 20210407-Screening rules management v01.00.
- [RD14] FX Screening Rules Answers.
- [RD15] ETIAS Risks Screening Business Processes (1).
- [RD16] 2020.05 COM study on opportunities and challenges for the use of AI in border control, migration and security.
- [RD17] AI in the OM of Large-scale IT Systems - eu-LISA report.
- [RD18] CRRS BUC List.
- [RD19] Annex 8.2 ETIAS Business Use Case Model 04 01 00.
- [RD20] Annex 8.3 ETIAS Common Objects 04 01 00.
- [RD21] ETIAS Risk Profiles v17 (ETIAS new mock-ups).
- [RD22] CRRS External ICD 03 00 00.
- [RD23] CRRS External ICD 03 00 00.
- [RD24] 02017R2226-20210803 EES REG.
- [RD25] 02019R0818-20210803 Interoperability LEA Immigration REG v1.
- [RD26] Annex 13.1 CRRS Technical specifications Requirements
- [RD27] Annex 13.2 CRRS Technical specifications Architecture building blocks
- [RD28] Annex 6.0 CRRS Requirements
- [RD29] EDPS Opinion on the European Commission’s White Paper on Artificial Intelligence
- [RD30] Declaration on ethics and data protection in AI

[11] Section 6 6 2 in “Annex 13 1 CRRS Technical specifications Requirements” [RD26] and Section 3 8 in “Annex 13 2 CRRS Technical specifications Architecture building blocks” [RD27] for specific Architecture Building Block (ABB) details Analytics and Machine Learning requirements for CRRS Tender Technical Specifications (TTS)

1.9 Definition of terms

The following table describes the terms defined, by the authors, used in the scope of this study and document.

Term	DESCRIPTION
Algorithm	In mathematics and computer science, an algorithm is a finite sequence of rigorous instructions, typically used to solve a class of specific problems or to perform a computation
Artificial Intelligence	Artificial intelligence (AI) is perceiving, synthesizing, and inferring information - demonstrated by machines, as opposed to intelligence displayed by animals and humans. Example tasks in which this is done include speech recognition, computer vision, translation between (natural) languages, as well as other mappings of inputs.
Correlation	In statistics, correlation or dependence is any statistical relationship, whether causal or not, between two random variables or bivariate data.
EU Information Systems	EU information systems, namely the Entry/Exit System (EES), the Visa Information System (VIS), the European Travel Information and Authorisation System (ETIAS), Eurodac, the Schengen Information System (SIS), and the European Criminal Records Information System for Third-Country Nationals (ECRIS-TCN). Additionally the interoperability components are: European search portal (ESP), shared biometric matching service (shared BMS), a common identity repository (CIR) and a multiple-identity detector (MID) should be established as interoperability components. [12]
Hit	Existence of a correspondence established by comparing the data recorded in an application file of the ETIAS Central System with the specific risk indicators referred to in Article 33 or with the personal data present in a record, file or alert registered in the ETIAS Central System, in another EU information system or database listed in Article 20(2) ('EU information systems'), in Europol data or in an Interpol database queried by the ETIAS Central System.[13]
Risk Indicator	A combination of data established by the ETIAS Central Unit that might drive to a risk condition related to security or illegal immigration or a high epidemic risk, including one or several of the following (acc. to Art. 33(4) of the ETIAS Regulation): (a) age range, sex, nationality; (b) country and city of residence; (c) level of education (primary, secondary, higher or none); (d) current occupation (job group)[14]
Risk Profile	A combination of specific risk indicators, screening rules and matching threshold associated with a certain weight that can produce a risk condition outcome related to security or illegal immigration or a high epidemic risk. [1]

[12] (From Interoperability Regulations (EU) 2019/817 and 2019/818)
[13] (From ETIAS Regulation (EU) 2018/1240, Art 3(14))
[14] (From ETIAS Regulation (EU) 2018/1240, Art 33(4))

Term	DESCRIPTION
Ex ante assessment	Process to define, establish, and assess risk indicators during profile implementation (equivalent to the development environment before commissioning the Profile on the production environment). The ex-ante assessment is based on the Bayes theorem that defines the matching rate given the estimated real probability, the reliability rate and the inclusion rate that are defined as following: - Inclusion rate = P(RISK| <Threshold) - Reliability rate = P(<Threshold | RISK) - Estimated real probability = P(RISK) - Matching rate = P(<Threshold) NOTE: Implements Articles 7(2)(c) and 33 of ETIAS Regulation (EU) 2018/1240.
Ex post assessment	Process to implement and evaluate risk indicators of a risk profile already in production environment, or in active state. NOTE: Implements Articles 7(2)(I) and 33 of ETIAS Regulation (EU) 2018/1240.
Screening Rules	The ETIAS screening rules shall be an algorithm enabling profiling as defined in point 4 of Article 4 of Regulation (EU) 2016/679 through the comparison in accordance with Article 20 of this Regulation of the data recorded in an application file of the ETIAS Central System with specific risk indicators established by the ETIAS Central Unit under paragraph 4 of this Article pointing to security, illegal immigration or high epidemic risks. [15]
Set of characteristics	Means distinguishing sets of observable qualities or properties identified based on information and statistics referred to in Article 33(2) of Regulation (EU) 2018/1240 and considering the data referred to in Article 33(4)(a) to (d) of that Regulation.
Specific group of travellers or Cluster	Means a group of third-country nationals of a known size and composition attributed sets of characteristics. [16]
Statistical data	Means the data, which is anonymised and used solely for the purpose of producing statistical reports. [17]
Statistical reports	Means an organised collection of statistical data, produced by the central repository in an automated manner according to a set of pre-established rules and stored in the central repository. [18]

Table 3. Definition of terms – Ontology

[15] (From ETIAS Regulation (EU) 2018/1240, Art 33(1))
[16] (From ETIAS CDD 2021/4981, Art 2(a))
[17] (From CRRS CDR (EU) 2021/2222, Art 1 (1))
[18] (From CRRS CDR (EU) 2021/2222, Art 1 (2))

2 Chapter 2 - Methodology

In this Chapter we provide a brief overview of the methodologies followed by the Consortium in order to identify and design the proposed Business Use Cases (BUCs). On this direction, analysis started with the literature processing methodology, illustrating the individual steps of data collection, reduction, presentation and verification. Subsequent to the literature processing methodology, the information gathering methodology provided all the necessary means to extract the required information from the stakeholders. Specifically, one-on-one interviews, focus group discussions, collaborative workshops, and surveys were considered. Upon all the necessary information was collected, the Consortium proceeded with the definition of a representative methodology for defining significant Business Use Cases (BUCs) for the organization. Brainstorming sessions with eu-LISA and the different stakeholders made it possible to identify potential Use Cases. These were presented to the contracting authority, submitted for consultation with stakeholders and those that were accepted were prioritized to be detailed and further elaborated.

The next step of the study included the identification the business and stakeholders requirements steaming from the agreed BUCs, along with a concrete methodology towards their elicitation. Key steps towards the elicitation methodology include their organization into requirements packages, the quality review, and the requirement management. After the identification of Business Use Cases and requirements analysis, the next step consisted in preparing a specific Chapter of the current study with a methodology for Identifying Artificial Intelligence tools. Specifically, relying on the *DMADV methodology (Define, Measure, Analyse, Design, Verify)* in order to support the definition of the Business Use Cases, along with their stakeholder and business requirements.

Appendix IV – Methodology for the identification of Business Use Cases, Requirements and AI Tools includes the broader description of the methodologies used in the current Study, including, the presentation of Business Use Cases, Requirements, along with the AI tools identification methodology.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[19] State of the art and beyond the state of the art techniques for statistical analysis will be considered

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

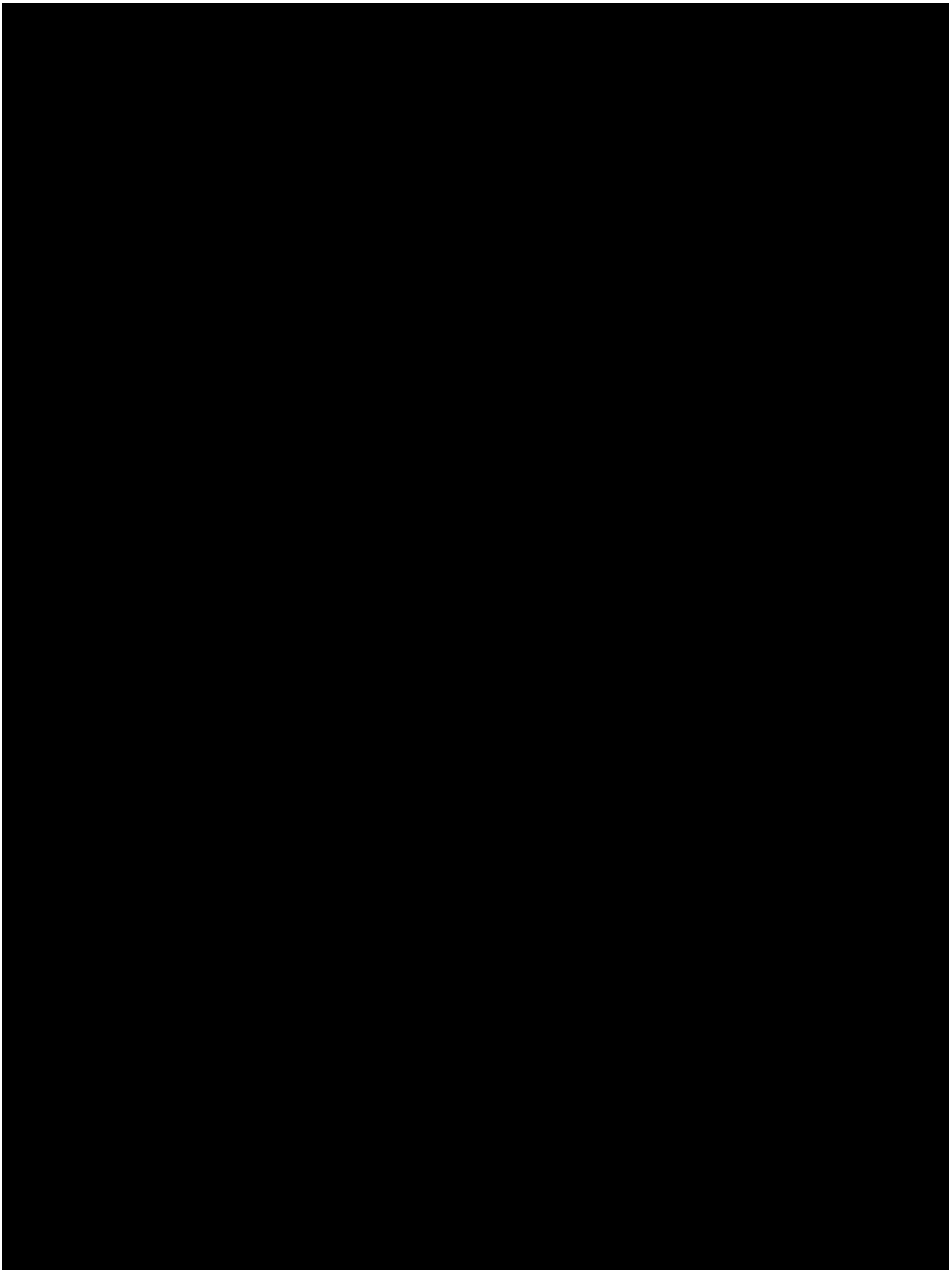
[redacted]

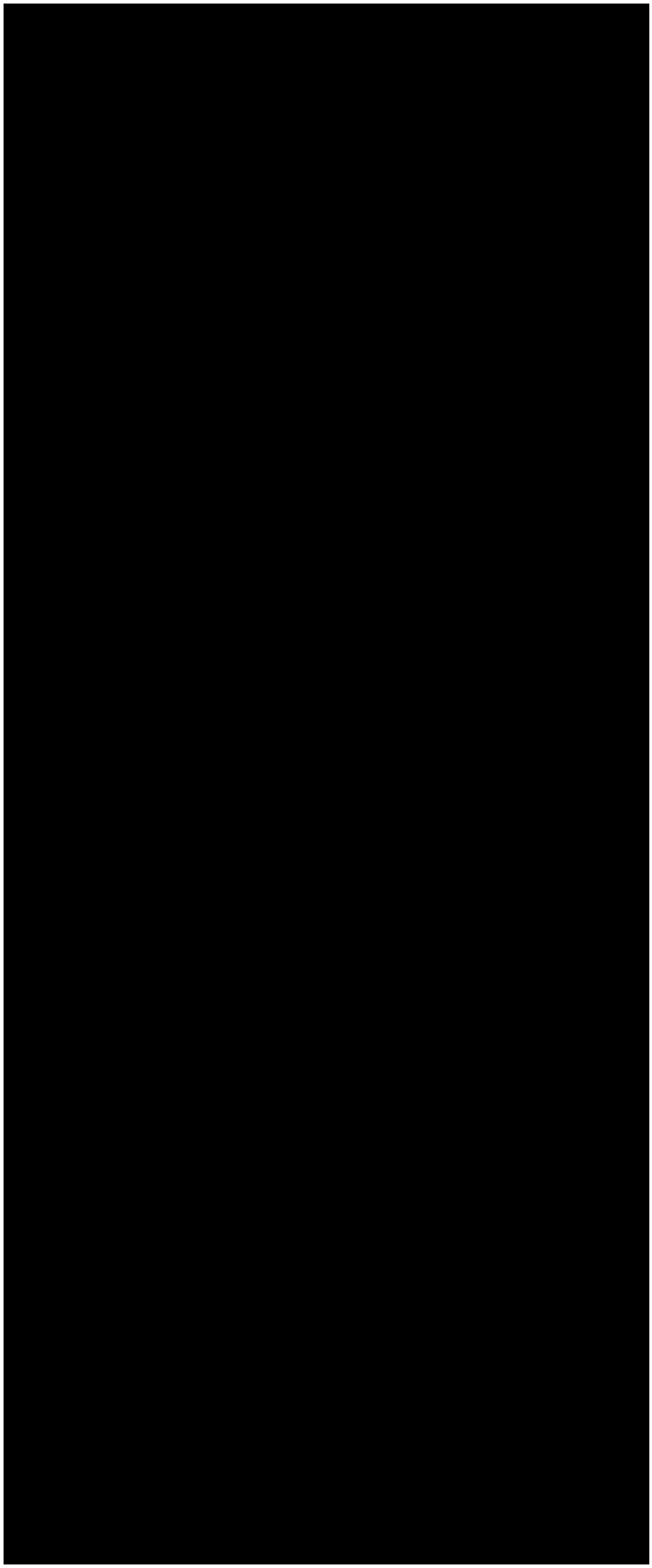
[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	
[redacted]	

[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]
[redacted]	

[redacted]





[redacted]

[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

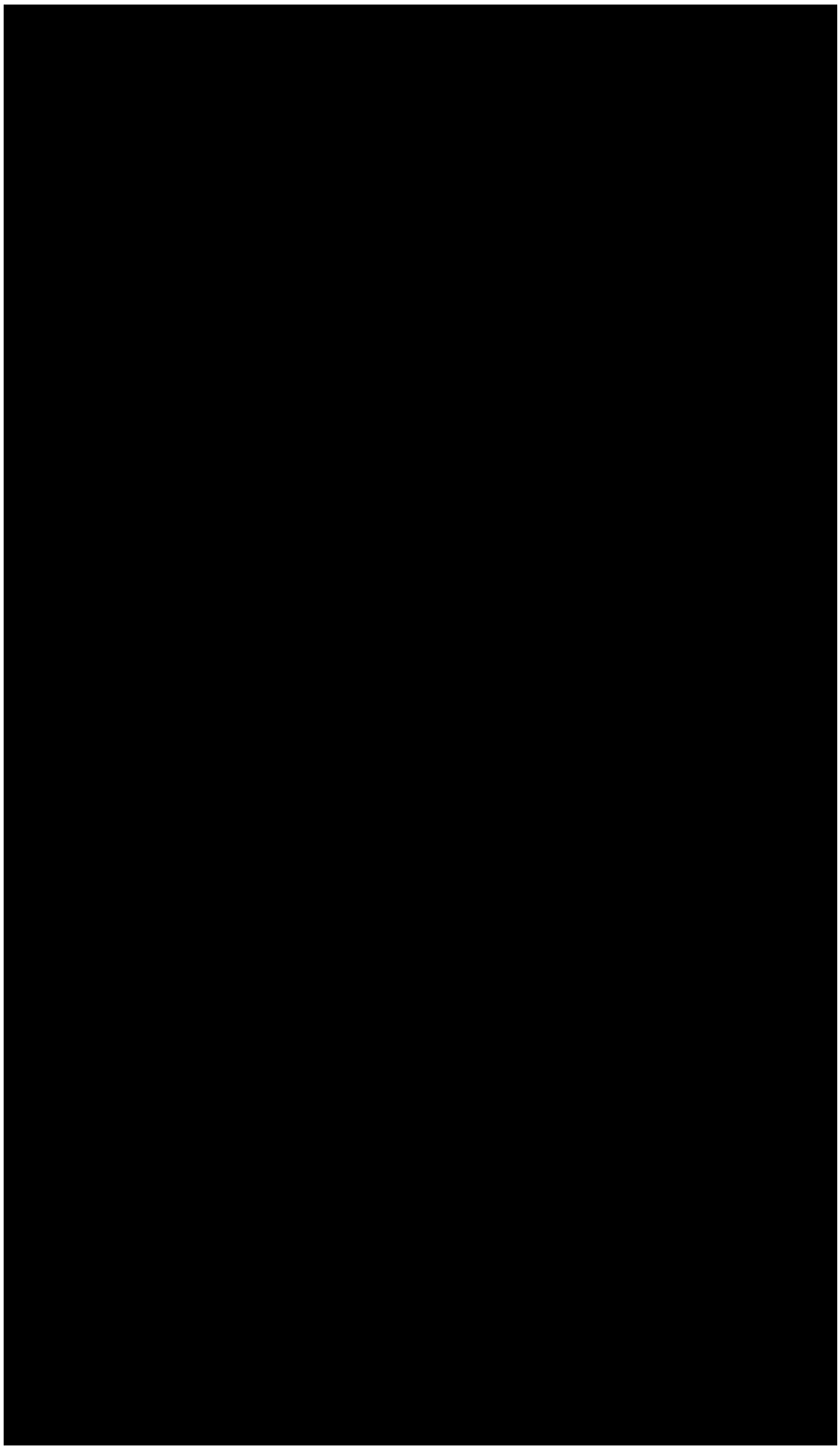
[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

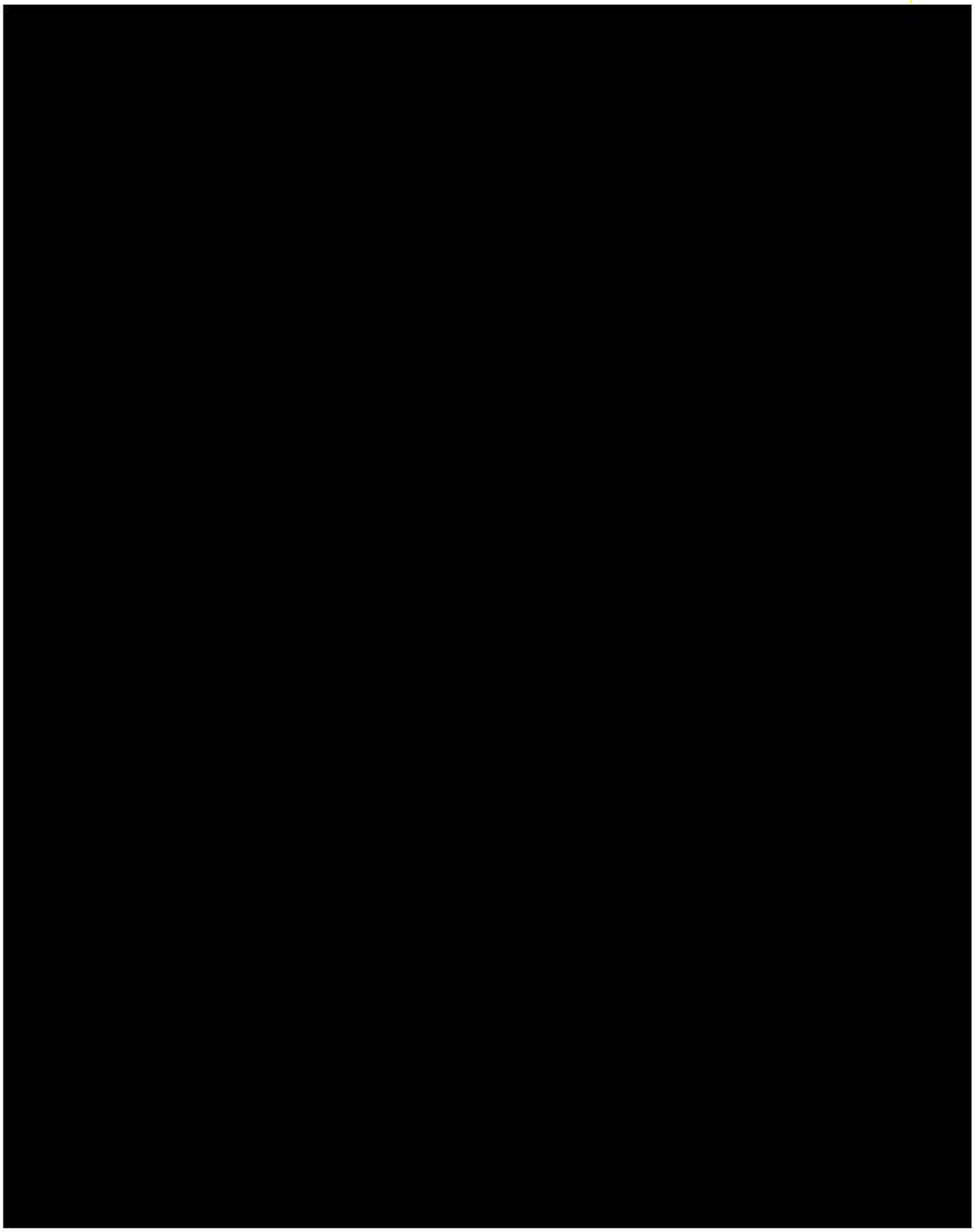


[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

[redacted]	[redacted]
[redacted]	
[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]
[redacted]	

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

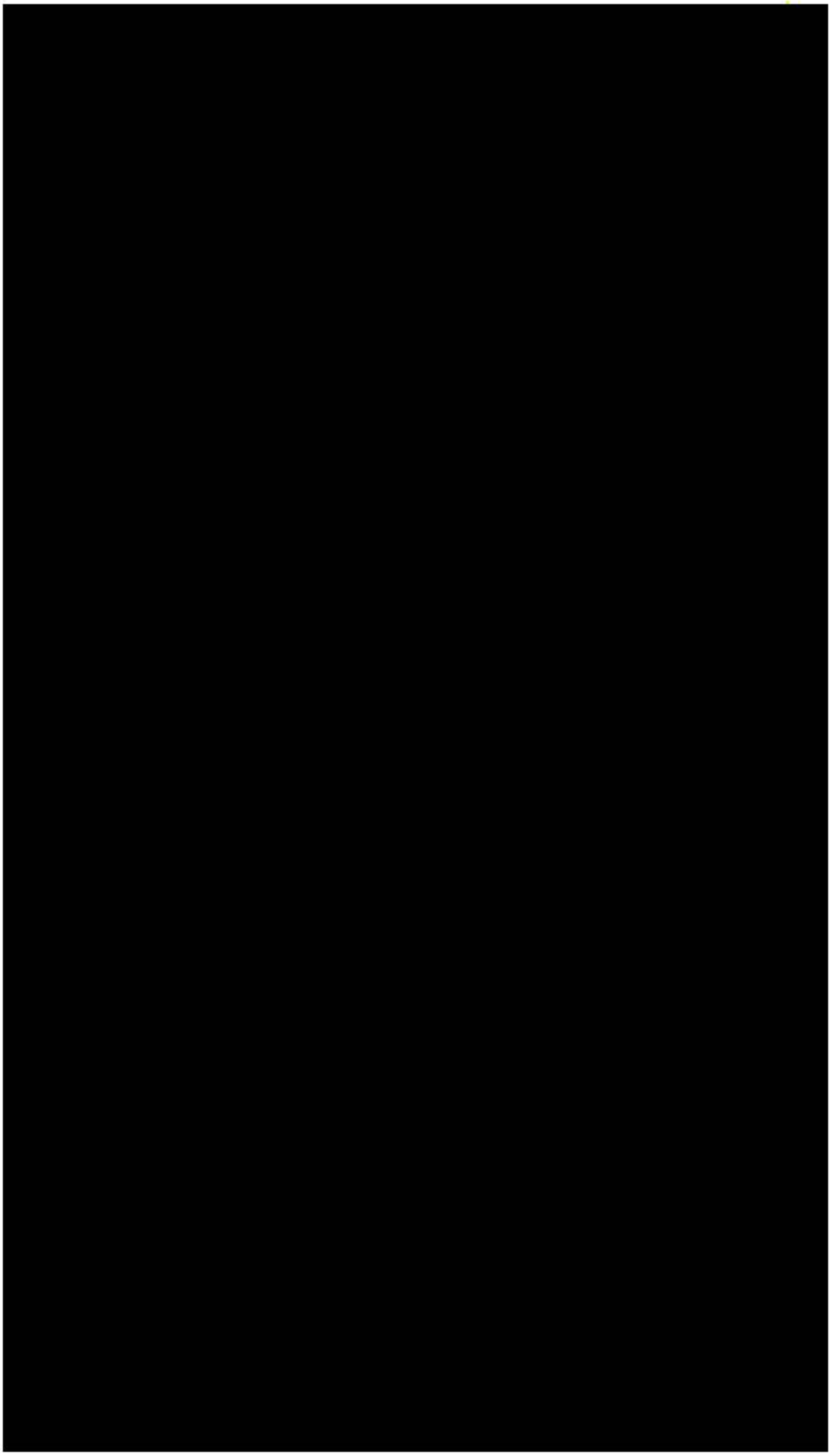
[redacted]

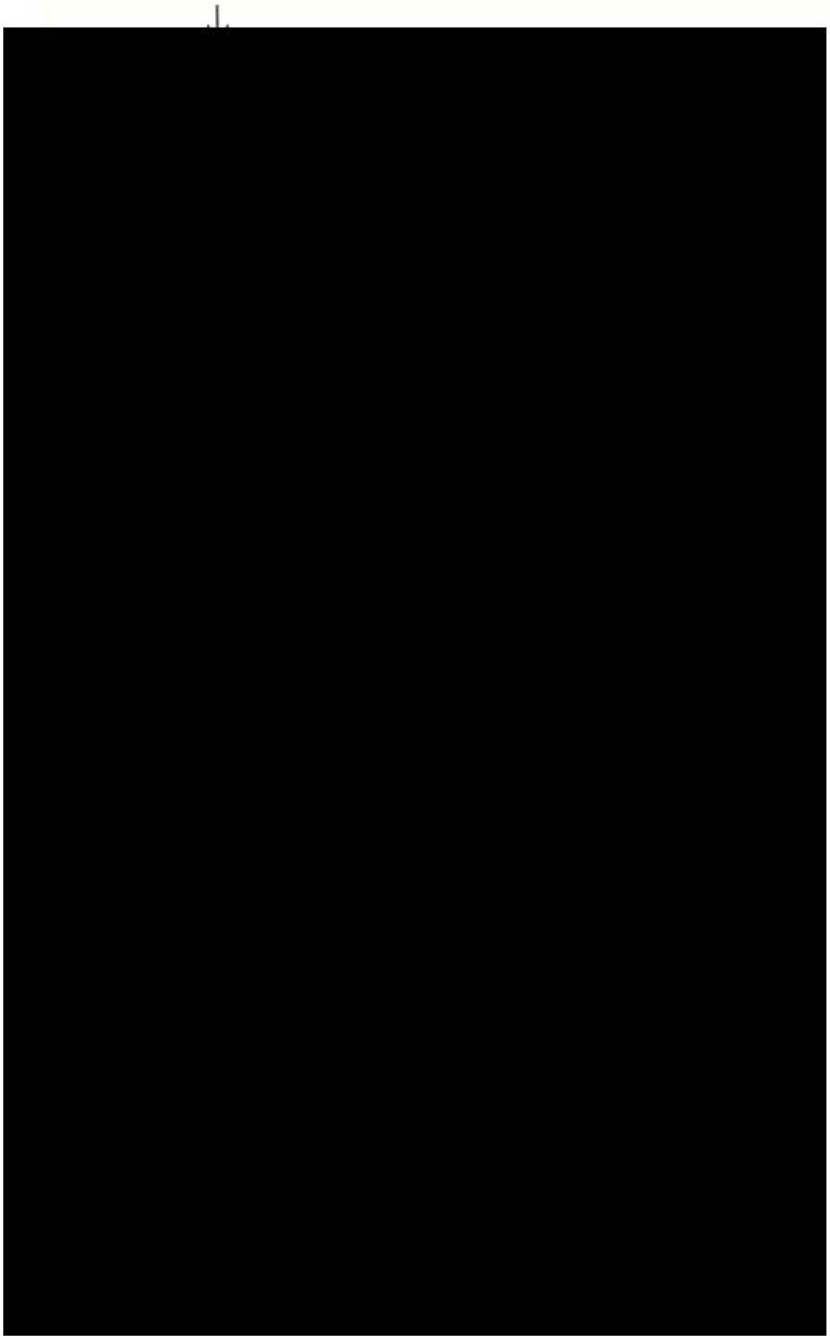
[redacted]	[redacted]
[redacted]	
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[20] State of the art and beyond the state of the art techniques for statistical analysis will be considered

[redacted]	[redacted]
[redacted]	
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]
[redacted]	

[redacted]





[redacted]

[redacted]

[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

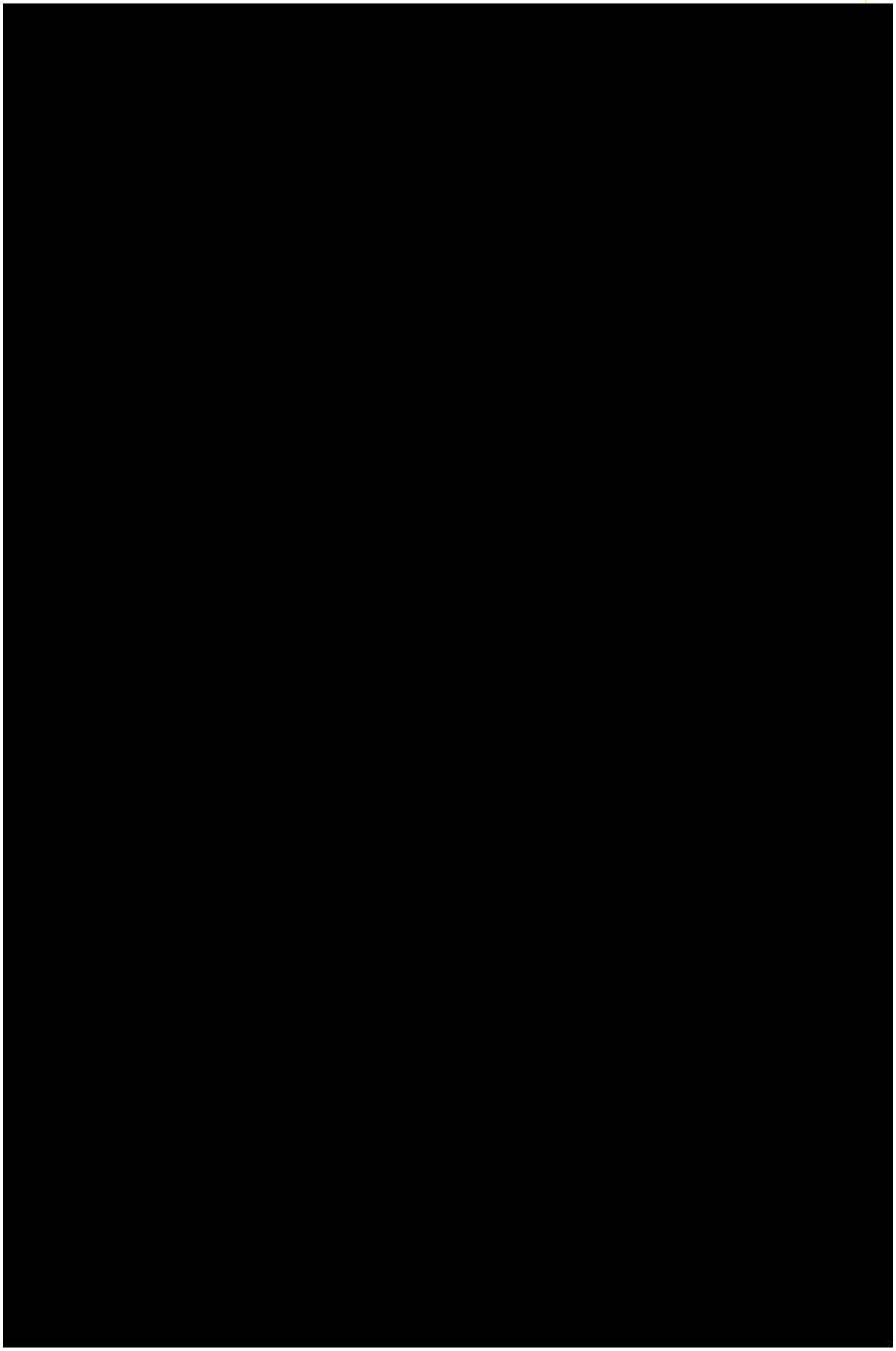
[redacted]

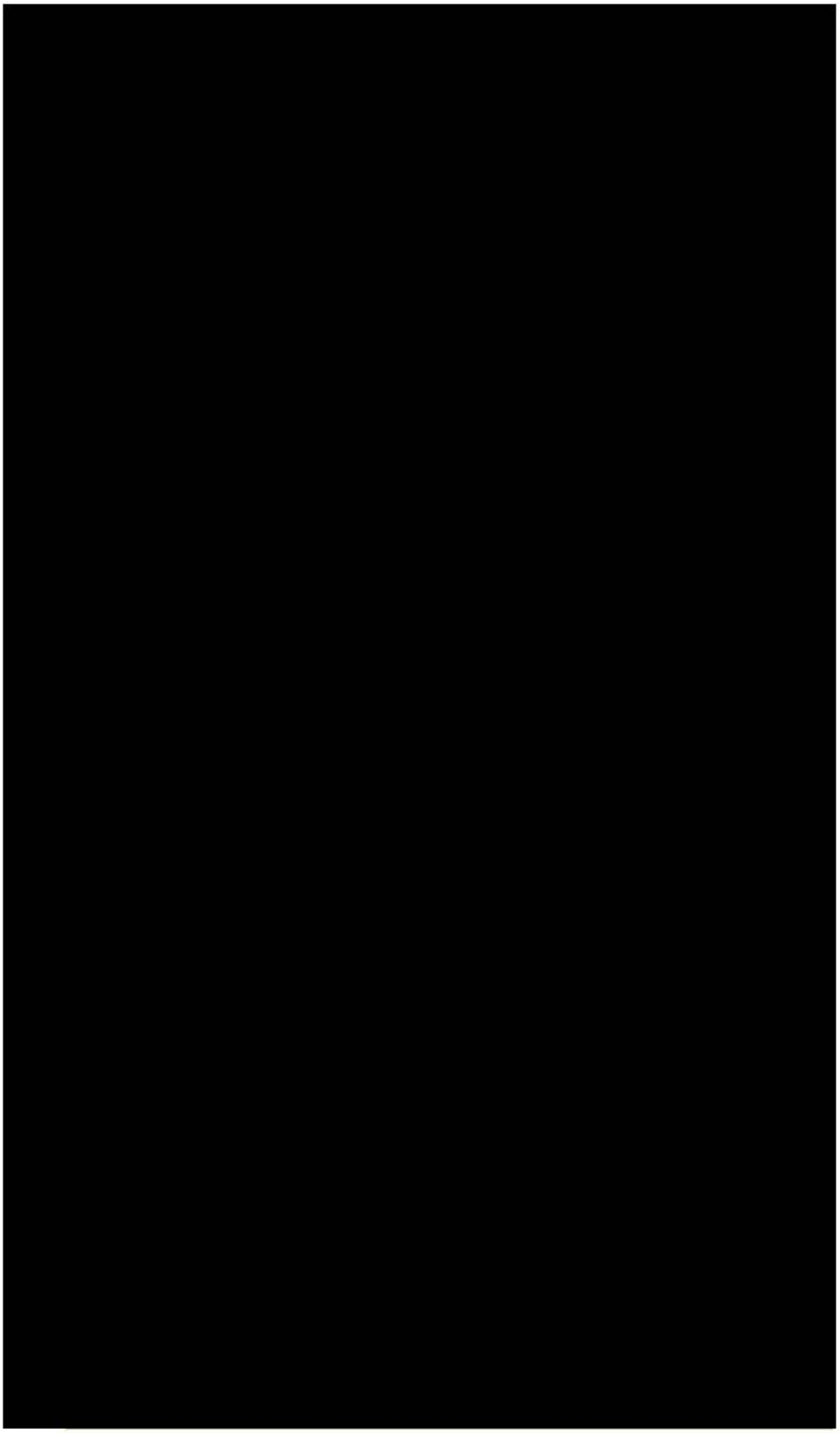
[redacted]	[redacted]
[redacted]	
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	
[redacted]	[redacted]
[redacted]	

[21] State of the art and beyond the state of the art techniques for statistical analysis will be considered

[redacted]	
[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]
[redacted]	

[redacted]





[redacted]

[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]

[redacted] [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

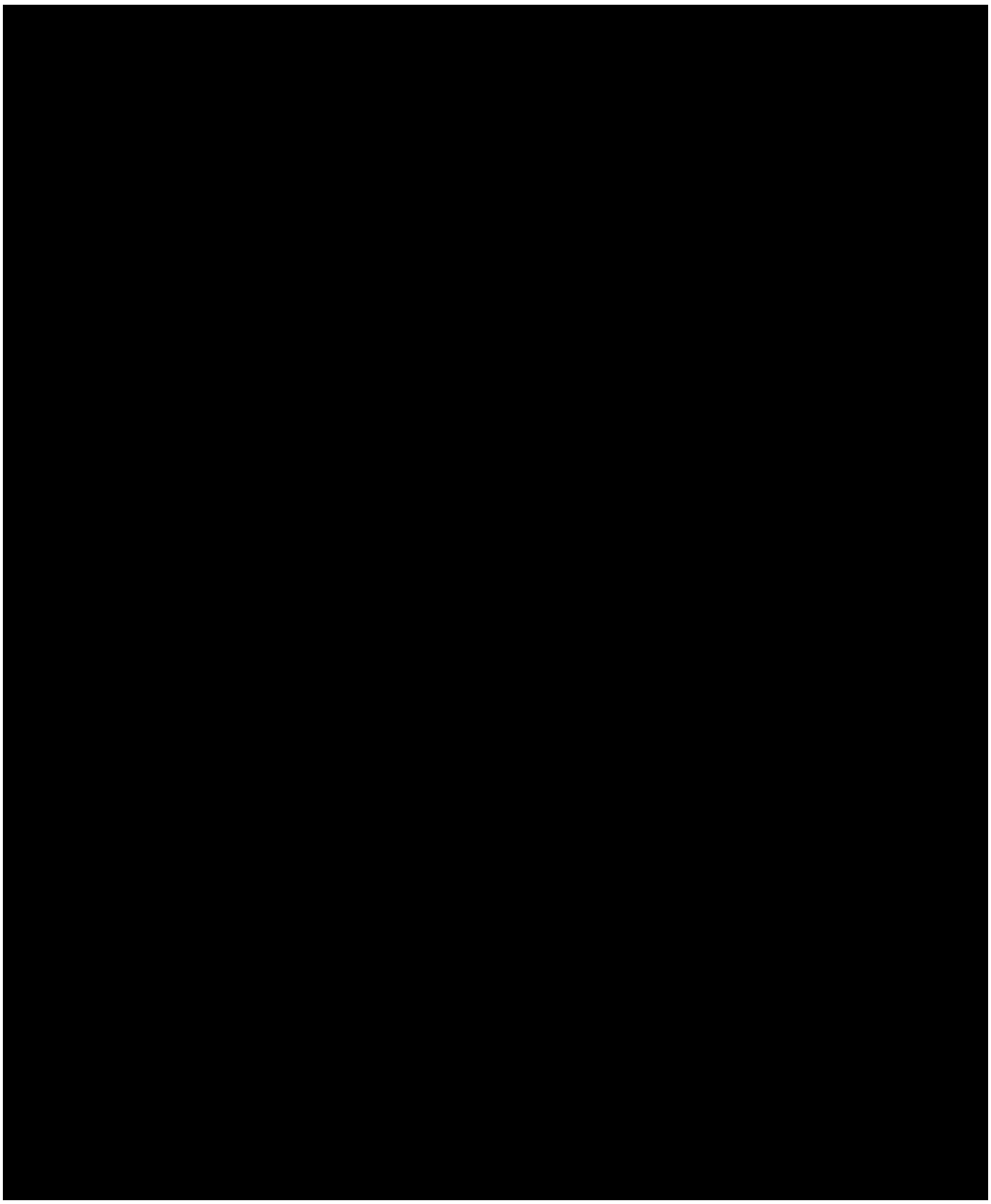
[redacted]

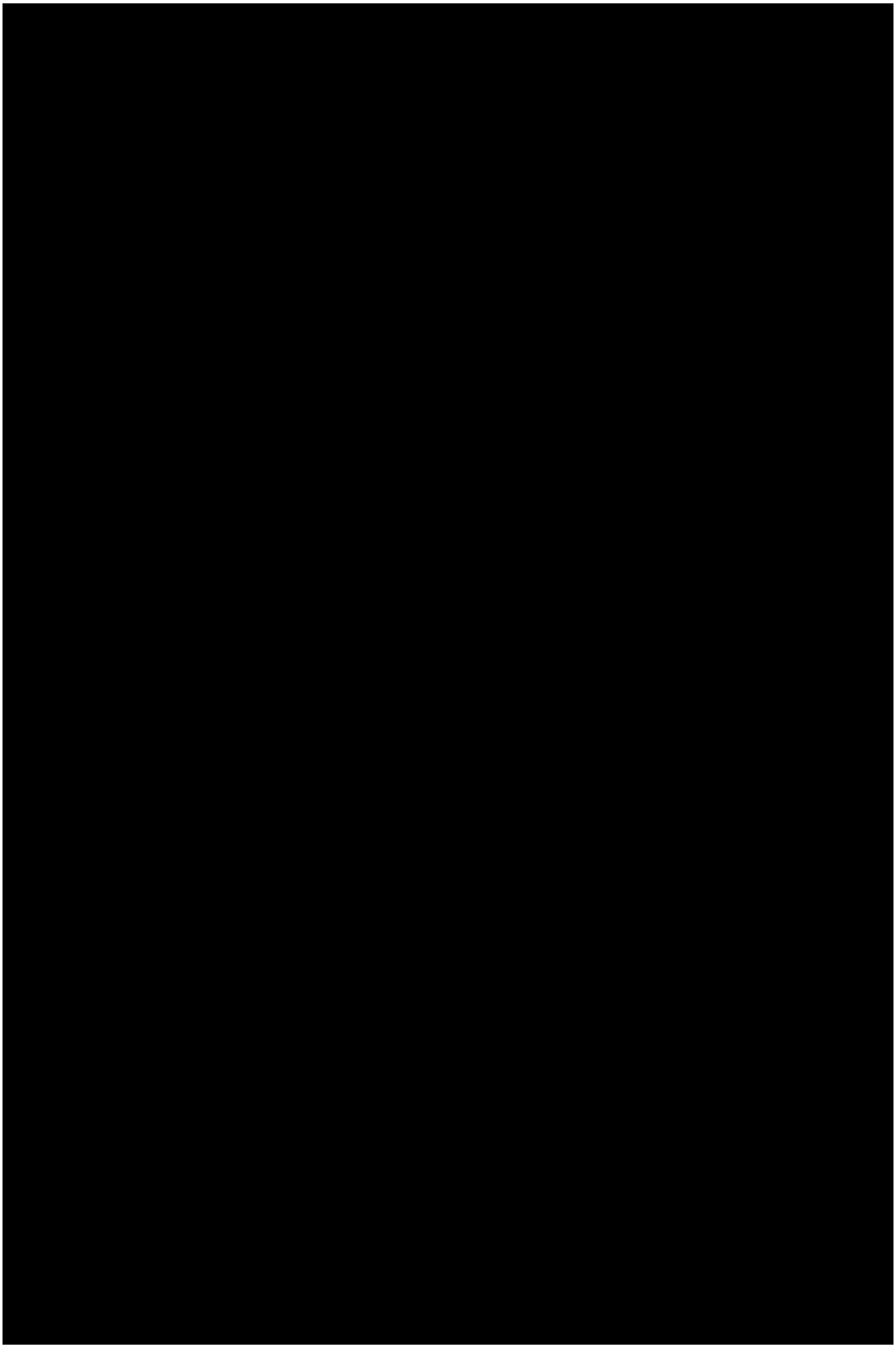
[redacted]

[redacted]	[redacted]
[redacted]	
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]
[redacted]	
[redacted]	
[redacted]	[redacted]

[redacted]





4 Chapter 4 - Requirements

4.1 RP.01 Identification of risk for a specific group of travellers.

4.1.1 Requirements Hierarchy (RP.01)

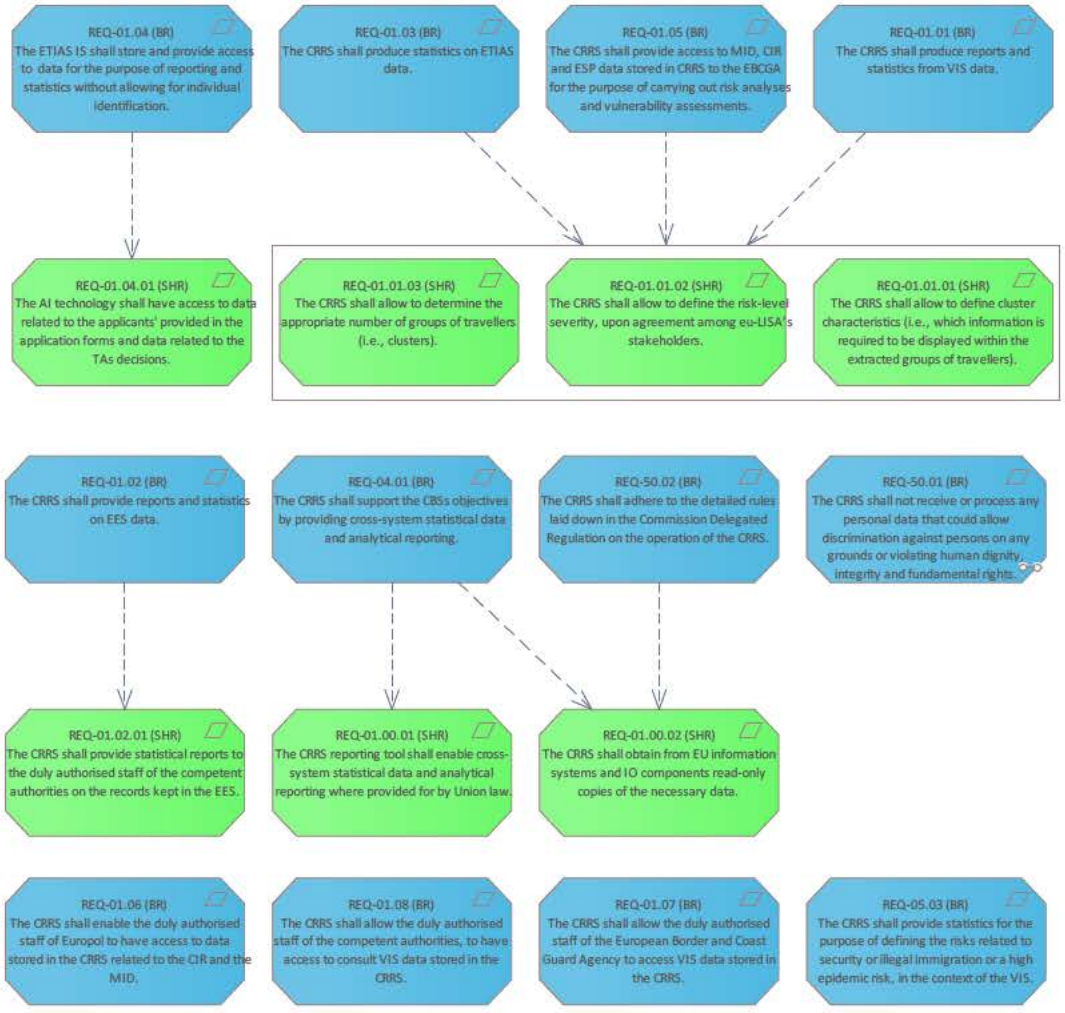


Figure 14 RP.01 Identification of risk for a specific group of travellers.

The complete definition of requirements is available later in Section 9.1.

4.2 RP.02 Review and validate identified risk profile.

4.2.1 Requirements Hierarchy (RP.02)

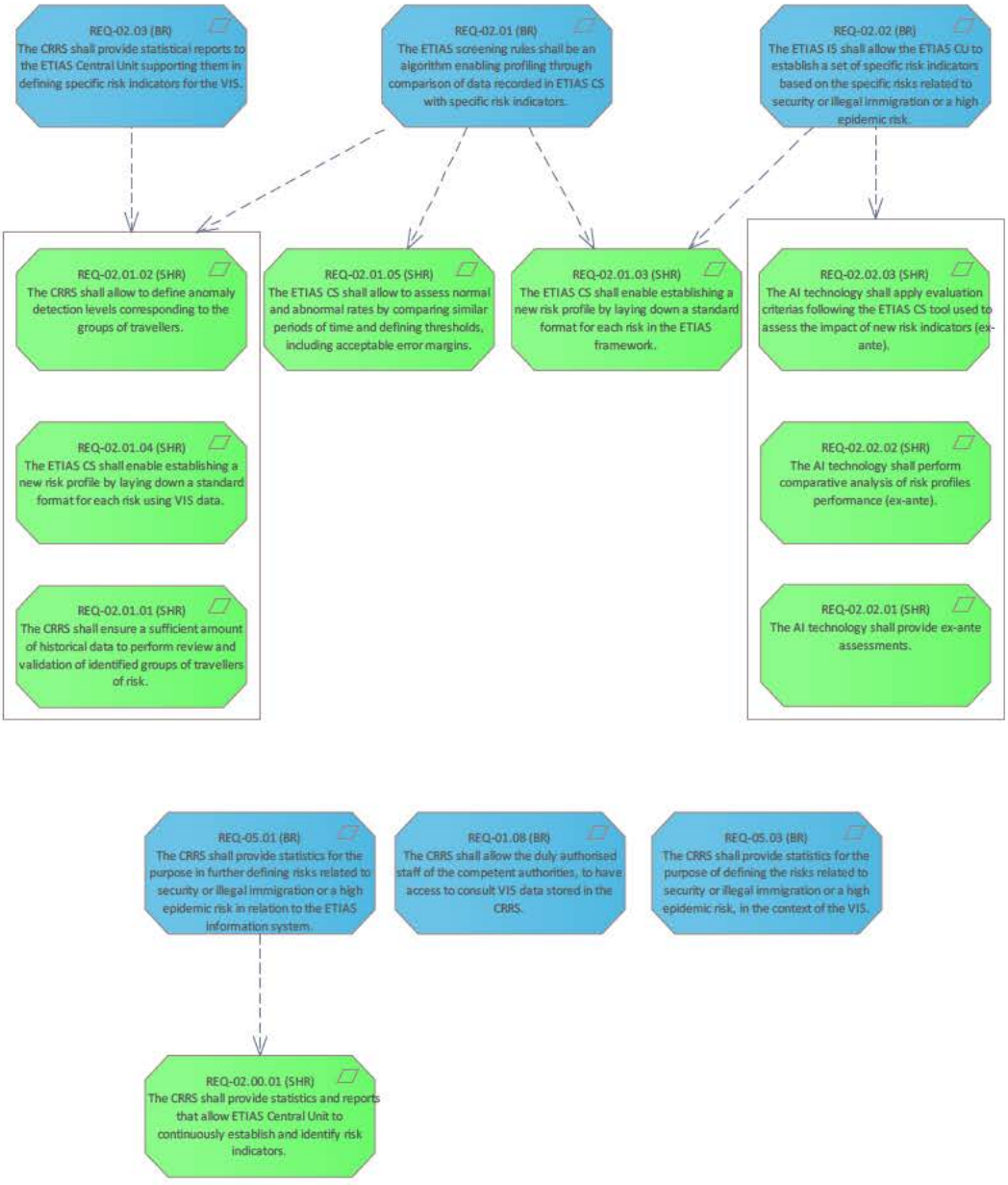


Figure 15 RP.02 Review and validate identified risk profile.

The complete definition of requirements is available later in Section 9.2.

4.3 RP.03 Analyse risk profiles correlations.

4.3.1 Requirements Hierarchy (RP.03)

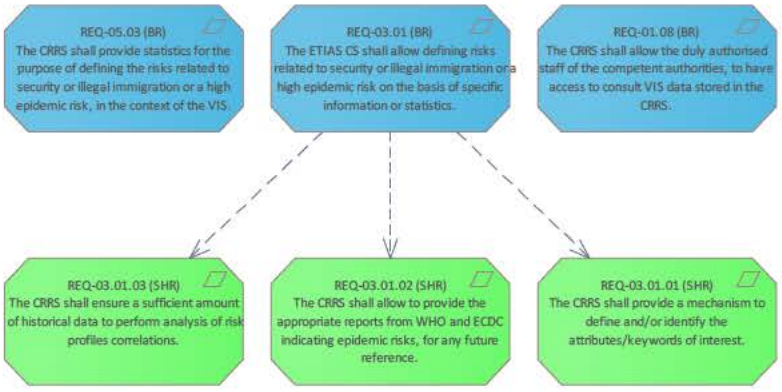


Figure 16. RP.03 Analyse risk profiles correlations.

The complete definition of requirements is available later in Section 9.3.

4.4 RP.04 Automatic processing of textual information from various sources

4.4.1 Requirements Hierarchy (RP.04)

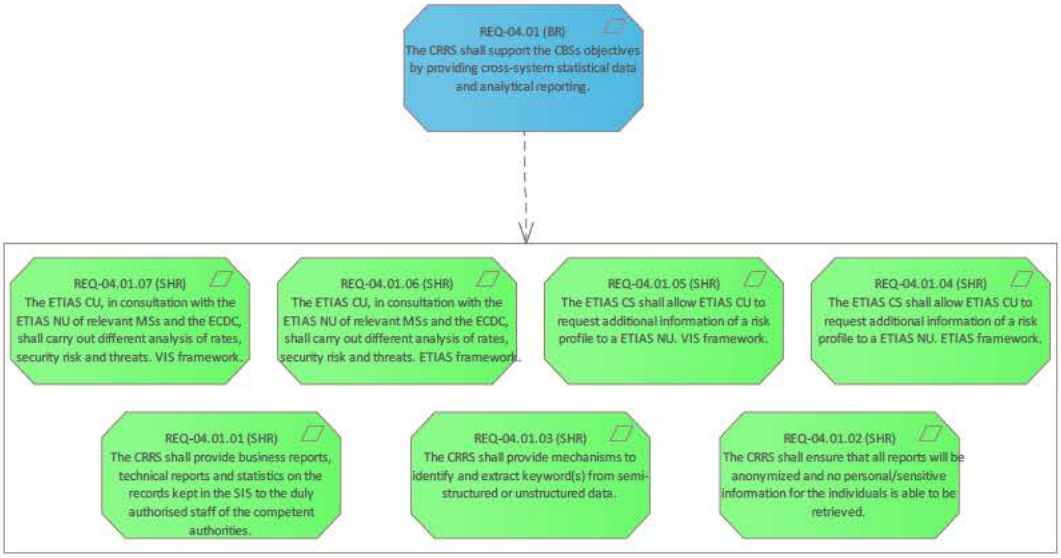


Figure 17. RP.04 Automatic processing of textual information from various sources.

The complete definition of requirements is available later in Section 9.4.

4.5 RP.05 Risk profile validity assessment.

4.5.1 Requirements Hierarchy (RP.05)

Figure 18. RP.05 Risk profile validity assessment.

The complete definition of requirements is available later in Section 9.5.

4.6 RP.06 Virtual assistance to provide guidance on certain activities.

4.6.1 Requirements Hierarchy (RP.06)

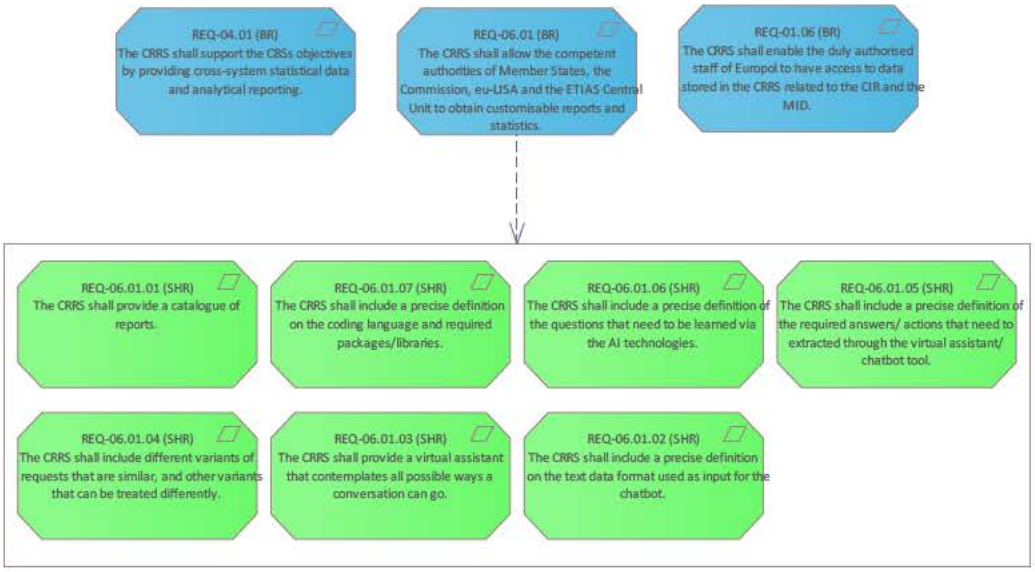


Figure 19. RP.06 Virtual assistance to provide guidance on certain activities.

The complete definition of requirements is available later in Section 9.6.

4.7 RP.07 Verification process when a hit takes place.

4.7.1 Requirements Hierarchy (RP.07)

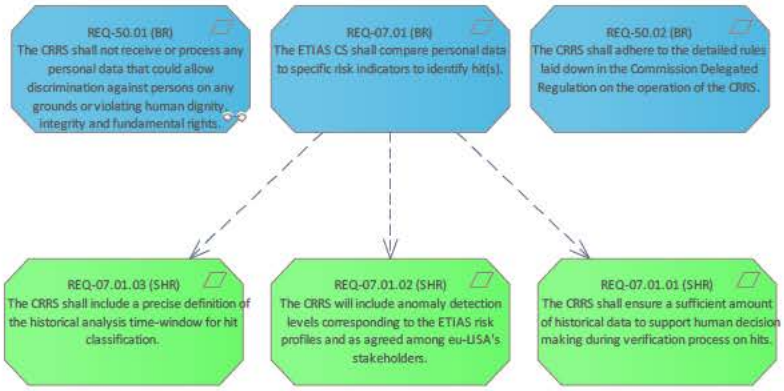


Figure 20. RP.07 Verification process when a hit takes place.

The complete definition of requirements is available later in Section 9.7.

4.8 RP.50 Reusable requirements.

4.8.1 Requirements Hierarchy (RP.50)

Figure 21. RP.50 Reusable requirements.

The complete definition of requirements is available later in Section 9.8.

5 Chapter 5 – Applicable AI Technologies & Assessment in CRRS

5.1 Overview of the Current Architectural Landscape

Before we deep dive into the analysis and assessment of the AI Technologies for the various Business Use Cases (BUCs), it is essential to emphasise that the current study is relied on the technical specifications and the current architectural landscape of CRRS[22]. CRRS accepts data from various external systems. The next figure depicts all data sources (i.e. different systems) that fill CRRS with data.

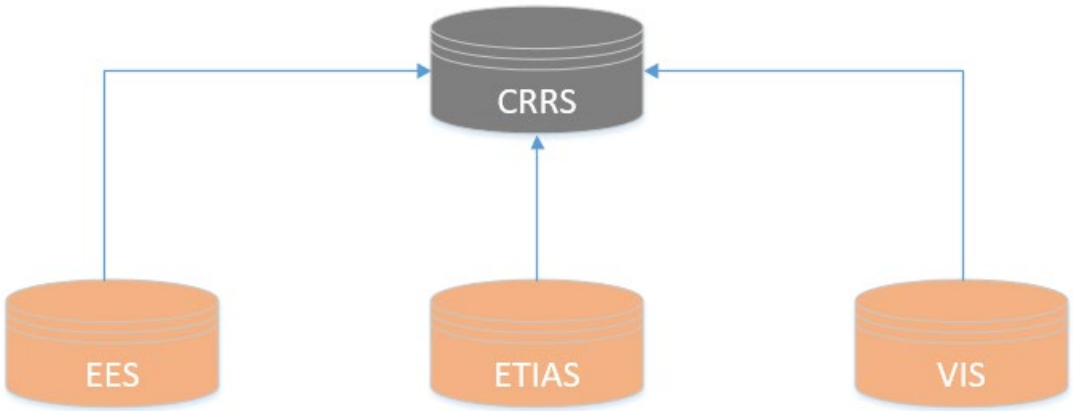


Figure 22. The various systems that provide data to CRRS

The next subsection (i.e., section 5.1.1) provides a small overview of the current data analytics processes, i.e., architectural landscape of CRRS.

5.1.1 CRRS Data Analytics Architectural View

Figure 29 depicts the holistic view of the CRRS data analytics process as it is designed until now. Data analytics is divided into two distinct components: (i) *CRRS Manual Data Analytics*, and (ii) *CRRS Predictive Analytics* (i.e., with the support of Machine Learning (ML)). Example of the data analytics usages include the identification and explanation of business data trends based on data-driven ML-based models, the establishment of DQ standards for achieving the objectives of the CBSs; the improvement of the various ETL process by fine-tuning Transformation Rules and CRRS data structures; and to further look for reasons regarding excessive storage consumption.

In both components, the procedure starts by ingesting data into CRRS from the various CBSs (i.e., VIS, SIS, ETIAS, EES, etc.) and a consequent step of Analysis is being implemented. Additionally, data is inserted to CRRS also from DQ Statistical Reports. During the Analysis step, three types of tasks occur:

[22] CRRS-E1-WP3-003 High Level Design

(i) Data structures management;
(ii) Determination of Quality Standards;
(iii) Description of the logic behind Data Transformations.

The next step of the diagram below illustrates the envisioned transformation procedure from the manual to the predictive data analytics processes (i.e., by involving Artificial Intelligence with human intervention). Once the environment is established, the process of producing results from machine learning algorithms is expected to examine the historical data, and further use them to generate a specified model and store the results and parameters into a data store. A further step towards the current architectural landscape, is to use new data measurements, as reaching CRRS, on the updated model, to generate new predictions and produce the corresponding results/ outcomes. The discussed architecture is depicted in detail in the following image:

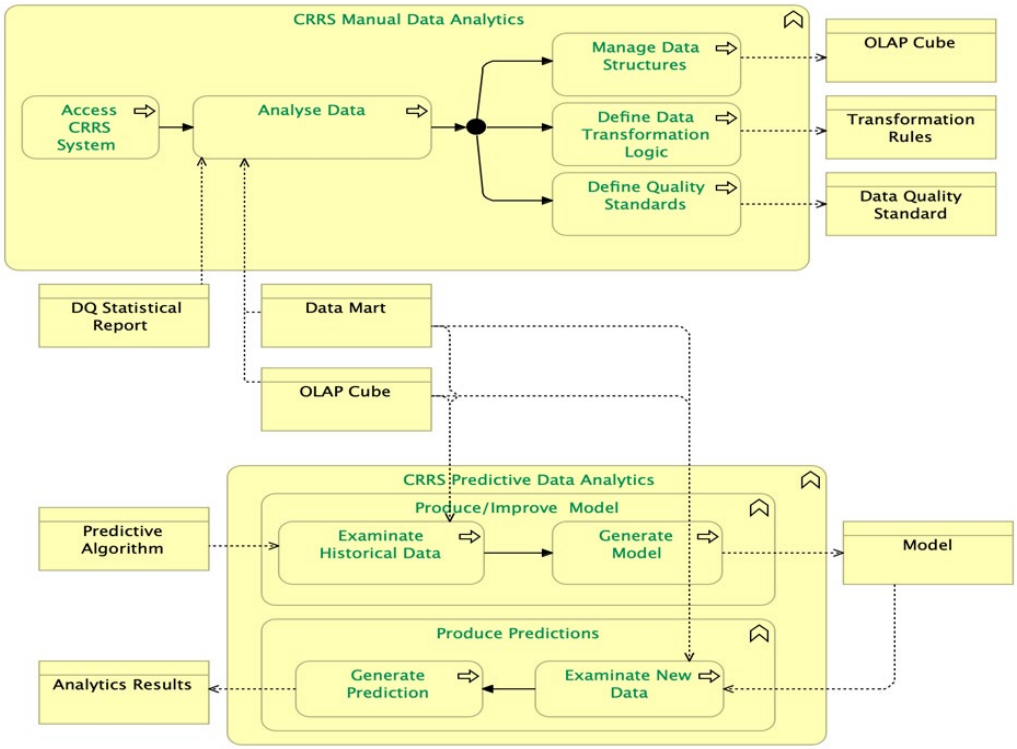


Figure 23. CRRS Data Analytics View

5.2 Methodology for Identifying Open-Source AI Tools

5.2.1 Analysis Step

In this step of the study, we proceed with the subsequent steps of the **(DM)ADV Six Sigma framework**. Specifically, the Business Use Cases (BUCs) that were gathered in previous phases (**Phase 1** and **Phase 2**) are analysed to proceed to the next step (**Design Step**) and suggest the suitable **AI technologies** and **tools** that will structure the envisioned technological landscape of eu-LISA. During this step, there is a discussion about the dimensions that apply within the terms of **CRISP-DM** (**Cross-Industry Process for Data Mining**) Methodology for the overall **AI assessment**, applying for each individual BUC; the concept of **Machine Learning Operations** (**MLOps**); as well as **Data Operations** (**DataOps**).

5.2.1.1 CRISP-DM Methodology

CRISP-DM (i.e., **Cross-Industry Process for Data Mining**) methodology provides a structured approach towards planning all necessary steps of a data analytics project. It is a robust and well-proven methodology, with powerful

practicality and flexibility to several business analytics problems, while the individual steps of the process (i.e., as explained within the following paragraphs) will be adjusted according to the needs of eu-LISA. It is crucial to highlight that CRISP-DM refers to a dynamic process that will continuously capture change requests coming from eu-LISA, either from the business or from the technical perspective. On this direction, the first steps of the process (i.e., business understanding and data understanding) require an exploratory analysis to have a thorough understanding of the problems we are trying to solve. The proposed methodology aims to cover the challenging scenarios of: unrefined data (i.e., raw data) is not present in CRRS from the various CBSs; or data integration is not built yet; or unified data models are not completely defined.

Specifically, CRISP-DM methodology is divided into the following main steps:

- **Business Understanding:** The first stage of the **CRISP-DM** process is to understand the project's objectives from the business perspective. The goal of this stage is to uncover key factors that could influence the outcome of the project. Key steps during this first stage are the following: (i) main objectives setup; (ii) production of a concrete project plan; and (iii) determination of key success indicators (KPIs), i.e., business criteria that justify the success of the project.
- **Data Understanding:** The second stage of the **CRISP-DM** process corresponds to concrete record and understanding of the data sources which are listed in the project's resources (i.e., data sources extracted from the various CBSs – VIS, SIS, ETIAS, EES, etc.). It is crucial to highlight that in the current study we design a Data Science and Machine Learning (DSML) platform, able to process any type of data that reach CRRS, or the corresponding CBSs, or even from external data sources (i.e., anonymized reports from authorized users). In the scenario where data cannot be extracted, the respective Data Science team, should receive information from third-party systems (i.e., various CBSs), regarding its nature and structure. Via this practice, the system will be ready and in-place to receive and further process data upon their availability. To summarize, data understanding consists of the following procedures: (i) Data Collection; (ii) Data Description; (iii) Data Exploration, and (iv) Data Quality Verification.
- **Data Preparation:** This step is responsible for preparing the data for the modelling/ training phase. It is important to emphasize our assumption that the proposed DSML platform will receive data from heterogeneous data sources (various CBSs), which is already in place. Additionally, we assume that the data model for the training, testing and validation part of the Machine Learning/ AI process is already in place. The key activities of this step are: (i) Data Selection, i.e., selecting the proper columns and rows for the analysis; (ii) Data Cleaning for model training, i.e., by ensuring that no-missing values are included and any other possible situations that might make the model training process to fail; (iii) Feature Engineering, by extracted representative features that facilitate the model training process; (iv) Data combination from two or more input datasets; (v) Data Formatting, i.e. converting categorical into numerical values, etc.
- **Modelling:** This step includes all necessary actions for developing the ML/ AI models, including the following key activities: (i) Model selection, which selects the best possible algorithmic formulation applying for different BUC; (ii) Test Design, corresponding to the practice that splits the input data into training, validation, and testing sets, and includes all actions for cross-validation – in our case the validation set corresponds to the upcoming (i.e. near-real time measurements that reach the CBSs); (iii) Model development, referring to fitting the selected model using the processed data from the previous steps; and (iv) Model assessment, by setting success indicators, including metrics, such as model's accuracy, validating the robustness of the model training process.
- **Evaluation:** This step differs from the model evaluation phase since it corresponds to the evaluation of the business indicators and requirements that were set during Phase 1 and Phase 2 of the project. The evaluation is made by answering to the following questions of: *how the business success criteria are met by using the already developed ML/AI models*; *if the AI/ML processes' outcomes are reasonable*; *whether all predefined steps of the project are correctly implemented and thus no critical steps have been excluded*; and finally, *the determination of the next steps, based on previous already achieved tasks*.
- **Deployment:** The last step of the current methodology corresponds to the delivery of the project's outcomes, and not only to the model deployment, which refers to ML/AI's deployment phase. Instead, in this step we can present, and/or deliver the results, to propose an efficient monitoring and maintenance plan of the current study applying on each BUC; and the final review of the end-to-end process, including any documentation, project summary report, presentations, etc. that could be extracted from the current

study.

All the above steps of the CRISP-DM methodology are concretely illustrated on **Figure 29.**

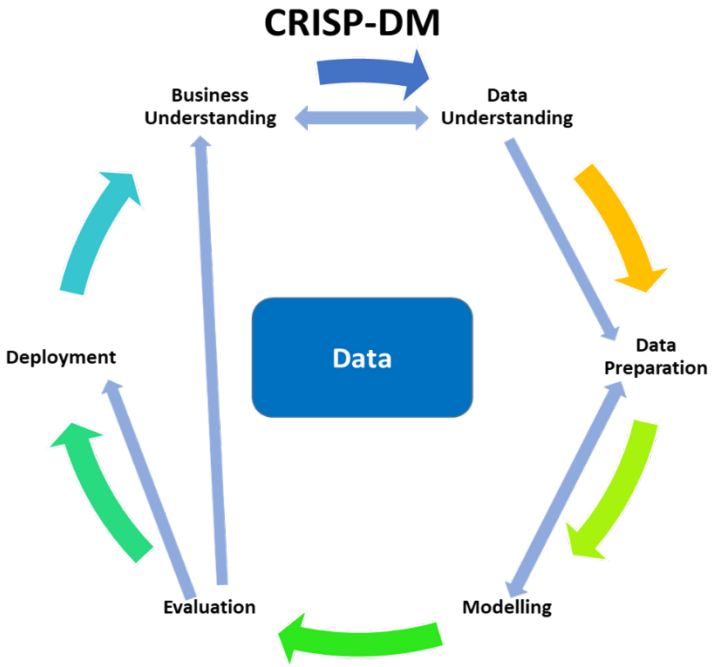


Figure 24. CRISP – DM Methodology

5.2.1.2 Machine Learning Operations (MLOps)

In general terms, **Machine Learning Operations** (**MLOps**) is considered a part of **DevOps** but dedicated explicitly to **Machine Learning (ML)** applications and is essential when it comes to machine learning processes that will be deployed in a production environment. **MLOps** is responsible for orchestrating correctly the corresponding processes and ensuring a reliable, effective and secure production environment. More specifically, the field of **MLOps** cover a wide range of tasks from the requirement elicitation to designing, testing and development. Hence, it is a practice for collaboration between professionals of different fields (data scientists, data analysts, data engineers etc.) to help manage production ML lifecycle. The key role of **MLOps** is to improve the quality of production ML and automate any engineering procedures. Specifically, **MLOps** refer to all engineering pieces that are gathered and enhance the deployment, run and train of AI/ML models. A standard MLOPs Process includes the following steps: *(i) Definition of the business needs, (ii) Data Pre-processing and Preparation, (iii) Model Training/ experimental phase, (iv) Model Validation, (v) Model Deployment, (vi) Model Monitoring, and (vii) Model Retraining.*

Business Needs Definition

Before starting with the actual steps of MLOps, one of the most challenging tasks is the clear definition of the business needs of the organization (including any stakeholders or external partners). It is essential to comprehend the exact needs of the organization and discover any places of improvement or further contribution. The business use cases that are created through this phase are the key to design a process that will cover any current or future needs and will enhance the existing procedure.

The Importance of Data Pre-processing for Model Training

The efficiency and accuracy of ML-based technologies in solving complex computational tasks without requiring severe human intervention are highly correlated with the availability and quality of the input data. However, in real

world scenarios, obtaining raw data that can be directly fed into any ML-based formulation is an extremely unlikely case. Usually, raw data, contain a significantly high percentage of missing or invalid information that can either collapse the ML model training, or completely mislead the learning tasks. For this purpose, data pre-selection, or data pre-processing are significant steps towards gaining valuable inferences from the unstructured and raw information. The main goal of data pre-processing is to find the most representative set of features that improve the performance of any subsequent ML process (i.e., classification/ clustering/ segmentation, etc.). It is used for converting complex data structures into concrete and representative attributes that can be efficiently modelled via any ML approach. Data pre-processing corresponds to the data cleansing procedures, i.e., missing values elimination or imputation according to certain rules, discretization of continuous attributes, binarization of attributes, and feature extraction, among others.

Model Training Phase/ Experimental Phase

Model training phase is the initial step of the **MLOps** life cycle. During this phase, data scientists test multiple different models, several hyper parameters, and multiple features. Key prerequisites for the model training phase include, code versioning, unit tests, scripts organized in a code template, containerized code, configuration files, requirement files, logs of results in a centralized and standardized way, and model registry, among others. Data scientists who are responsible for model training should already be granted with environments that adapt to their needs upon request on the customized infrastructures. Model training results will be systematically logged and centralized to ease the quick and collaborative iterations, while models will be versioned, since the above-mentioned steps are able to speed up the training/ experimental phase.

Model Validation

Model validation corresponds to the process that ensures the correct functionality of the trained model when it goes live into the production. This process includes three main aspects: (i) the raw performance of the model (i.e. accuracy, precision, ROC, and other evaluation metrics), (ii) meta-performance of the model, in terms of inference time, CPU, GPU needs, RAM, (iii) and Ethical considerations correlated with Artificial Intelligence, providing certain criteria that verify the absence of biases in the model predictions.

Implementation and Evaluation Metrics

The performance of several ML/ DL techniques is evaluated in terms of the accuracy, and the reconstruction errors of the models' loss function. Model's Accuracy is determined in the light of true positive T_p (i.e., genuine measurements correctly classified), true negative T_n (i.e., fake measurements correctly classified), false negative F_n (i.e., fake measurements incorrectly classified) and false positive F_p (i.e., genuine measurements incorrectly classified). It is defined as the ratio between the true outcomes towards the total number of outcomes, formulated as:

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

Additionally, the Area under the Receiver Operating Characteristic (ROC-AUC) score stands as an evaluation metric able to determine the degree of separability among the different categories, since it measures the classification performance of each model, per each classification class.

ROC curve illustrates the ratio between the True Positive Rate (TPR) i.e. $TPR = \frac{T_n}{T_n + F_p}$, and the False Positive Rate (FPR) i.e. $FPR = \frac{F_p}{F_p + T_n}$. Using the ROC curve, we evaluate how the model works while distinguishing between the variant <Event> categories. ROC-AUC scores that are close to 1.0 indicate highly robust models, which can perfectly determine the variant different classes. Moreover, we exploit the Precision, Recall, and F1-score metrics defined as:

$$Precision = \frac{T_P}{T_P + F_p}, Recall = \frac{T_P}{T_P + F_n}, F1 - score = \frac{2T_P}{2T_P + F_p + F_n}$$

High score on Precision metric indicates a lower False Positive Rate, i.e., achieving less fault-free data that were incorrectly marked as faulty. On the other hand, high score on the Recall metric demonstrates low ratio of False Negatives. Finally, F1-score provides the harmonic mean of Precision and Recall, by capturing these two measures

on a single metric. Concluding other evaluation metrics that can be used in the current study, are cross-entropy metrics, which vary with respect to the nature of each BUC.

Model Deployment

Model deployment is one of the most important/crucial steps in Machine Learning/ Deep Learning. Key prerequisite for a proper model deployment phase is the early definition of the requirements and specifications regarding the deployment process, including the model size, CPU/GPU and memory requirements, among others. The model deployment process will exploit the relevant inference cluster services of the proposed AI solution (as described explicitly on section 5.3.1) by complying with standard scalability, and security criteria.

Model Monitoring

Model monitoring corresponds to the process that ensures the model maintenance in operational condition by verifying that input data, and the predictions extracted by the model or even meta-performances of the model, which are faithful to what has been recorded during the model training phase.

Model Retraining

Since model's performances naturally decay over time, model retraining is a crucial step of the **MLOps** lifecycle. The trigger of the model retraining can be either *manual*, i.e., requested by eu-LISA's stakeholders or data scientists; *event-driven*, i.e., when new data reach the CBSs and thus is available for model training; and last, *automatically* when model performances are decreasing. An appropriate model-retraining pipeline is a necessary step on this procedure; however, this step requires human validation in order to evaluate the results of the model training phase, and act accordingly. It is crucial to emphasize that the results of the evaluation process correspond only to the models accuracy and evaluation metrics, which are responsible for providing high-quality Machine Learning outcomes. As part of the MLOPs and CRISP-DM

In eu-LISA's approach **MLOps** is a quality that should be taken into thoughtful consideration since every AI/ML model and its respective process will be deployed in a production environment that shall remain imperishable and stable.

5.2.2 Design Step

Within the current step of the (**DM)ADV Six Sigma framework**, we perform an assessment on the AI and ML applications that are of particular interest for eu-LISA, and mainly CRRS. Specifically, we separate our study into four main pillars: (i) the general foundation of **AI for decision making**; (ii) **Clustering analysis**, i.e., grouping of travellers into significant categories (i.e., clusters); (iii) **anomaly detection techniques**, for detecting patterns on the input data that significantly deviate from their expected behaviours; and (iv) **NLP-based approaches**, enabled by **conversational agents** and **virtual assistants**. In the following paragraph we describe in high detail the above-mentioned pillars, and thus we demonstrate the most appropriate and dominant state-of-the-art approaches that are presented in the recent literature. On this direction, we consider predictive analytics and time-series analysis techniques to capture interesting trends on CRRS's data distributions. Additionally, advanced analytics techniques, including Deep Neural Networks (DNNs) on the fields of NLP, anomaly detection, text and data classification are considered in the BUCs which are applicable. Key prerequisite to the current Design step, is the consideration of the key principles for Data Operations (DataOps) and Self-Service Analytics. For this purpose, a brief summary of DataOps is described on the following paragraph.

5.2.2.1 Data Operations (DataOps) & Self-Service Analytics

While section **5.2.1.2** was dedicated to **MLOps** and how it can be utilized to establish an AI-related environment, in the current paragraphs we present and describe Data Operations (**DataOps**). **DataOps** serves the same functionalities as **MLOps**, but still, they have significant differences. More specifically, **DataOps** corresponds to a collection of data analytics technical practices, workflows, and architectural patterns that affects all data engineering aspects and any procedure that is **data related**.

Via the adoption of a **DataOps** process, **eu-LISA** is able to achieve: (i) rapid innovation and experimentation; (ii)

extremely high-quality and very low error rates on the data-related processes; (iii) efficient and simplified collaboration across the different actors within eu-LISA by enforcing data protection rules; (iv) elimination of the manual work, and (v) clear measurement, monitoring and transparency of the data-related results. Precisely, **DataOps** shall be applied in all areas of the data architecture and aims towards a **more automated** and **agile-oriented data analytics landscape**. On the other hand, **Self-service Analytics** shall be considered part of the **DataOps** or at the very least close to **DataOps**. As Self-service Analytics are considered all processes and procedures that are enabled to transform the way data applications work and shall facilitate eu-LISA to become more **data-driven**.

5.2.2.2 AI Applications relevant to eu-LISA

AI for decision making

Nowadays, organisations must bring together increasingly complex techniques to make immediate decisions leveraging enormous amount of data. To improve the outcome of the decision models and incorporate several uncertainty factors, organisations include **AI techniques** for achieving maximum business differentiation. On this direction, decision making and intelligence is proved to promote the sustainability of cross-organizational decisions by building models using principles aimed at enhancing their traceability, replicability, and trustworthiness. Generally, **Decision Support Systems (DSS)**[23] consist of input, processing, and output components that reflects the decision-making process. Such systems are applicable to broader types of technology, including Business Intelligence (BI) and Data Analytics. In the field of Data Analytics, analysts deal with enormous amount of data in their daily basis. Consequently, AI techniques are often presented as an excellent choice for representing and solving such complex problems. The combination of decision support approaches and AI technologies results into the **Intelligent Decision Support Systems (IDSS)**[24], presenting several of the following capabilities:

- Learn or understand from experience (i.e., historical observations/ data);
- Clarify ambiguous decisions/ messages;
- Confront perplexing situations, by responding quickly and successfully to a new situation;
- Use reasoning in problem solving;
- Understand and infer in rational ways;
- Apply knowledge to understand the environment;
- Recognize the relative importance of the different attributes/elements that exist on a given situation.

A characteristic example illustrating the breadth of topics and variety of the AI tools utilised for decision making in EU's Law Enforcement and Criminal Justice[25] can be summarized on **Table 17**.

[23] https //www worldscientific com/doi/abs/10 1142/S0218213012400052
[24] https //www researchgate net/publication/321600307_Intelligent_Decision_Making_An_AI-Based_Approach
[25] https //www europarl europa eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN pdf

Application	Description	Reference
Predictive Policing	Algorithmic processes of datasets to reveal patterns of probable future offending & victimization. Predictive Policing may be divided into: • Methods aiming at predicting crimes or forecasting places and times with an increased risk of crime. • Methods aiming at predicting offenders or identifying individuals at risk of offending (or reoffending) in the future. • Methods aiming at predicting, or creating profiles like those of past offenders, and methods aiming at predicting victims of crimes, used to identify groups or individuals who are likely to become victims of crime.	Chapter 3; Section 3.1 of: "*Artificial Intelligence and Law Enforcement; Impact on Fundamental Rights*"; Policy Department for Citizen's and Constitutional Affairs; Directorate-General for Internal Policies; PE 656.295 July 2020;
AI and Criminal Justice	AI is explored/ used in European legal systems for several purposes: • Facilitating access to law (i.e., through chatbots). • Supporting alternative dispute settlement measures in civil matters, online dispute resolution, or "judge" profiling.	Chapter 3; Section 3.3 of: "*Artificial Intelligence and Law Enforcement; Impact on Fundamental Rights*"; Policy Department for Citizen's and Constitutional Affairs; Directorate-General for Internal Policies; PE 656.295 July 2020;
AI and Borders	Automated decision systems able to: • Classify individuals (e. g. as high risk, or high priority). • Generate scores, probability assessments, & other indicators supporting human decision-flag review or investigation. • Provide recommendations on applications, or even that render full decisions	Chapter 3; Section 3.4 of: "*Artificial Intelligence and Law Enforcement; Impact on Fundamental Rights*"; Policy Department for Citizen's and Constitutional Affairs; Directorate-General for Internal Policies; PE 656.295 July 2020;

Table 17. AI-enabled Intelligence Decision Support Systems used by European Organisations

Clustering Analysis

A cluster is considered as the collection of objects (i.e., groups) which present "**similarities**" among them, while they are "dissimilar" to the objects that belong to other clusters (i.e., groups). On this direction, clustering algorithms [26]aim to uncover the natural structures or relationships among the unlabelled data. **Clustering** is widely exploited Pattern Recognition approach since it is utilized for several decades in several fields of industry. For this purpose, several categories of clustering algorithms have been investigated[27]. Clustering-based approaches may be discriminated into **hierarchical** and **partitioning**, as it is illustrated on **Figure 30** [28]. Hierarchical clustering algorithms are usually based on the union between the two nearest clusters, while the beginning condition is realized by setting

[26] https //iajit org/portal/PDF/vol 5,no 3/15-191 pdf
[27] https //scholarsmine mst edu/cgi/viewcontent cgi?article=1763&context=ele_comeng_facwork
[28] http //hanj cs illinois edu/cs412/bk3/10 pdf

every data observation as a cluster. After a few iterations, hierarchical algorithms reach the desired number of clusters. Hierarchical algorithms may be divided into two categories: Agglomerative, and Divisive. Agglomerative algorithms focus on a bottom-up approach, by starting with many small clusters and then by merging them together to create bigger clusters. Divisive algorithms adhere to a top-down approach, by starting with a single cluster, and then break it up into smaller clusters. On the other hand, partitioning clustering algorithms classify the input information into multiple groups (i.e., clusters) based on the characteristics and similarity of the data, while the precise number of clusters that needs to be generated is specified by data analysts. In the partitioning methods, let (D) a database that contains (N) objects, then, the partitioning method constructs user-specified (K) partitions of the data in which each partition represents a cluster and a particular region. Partitioning-based algorithms can be discriminated into: distance-based, model-based, and density-based. On this direction, model-based and density-based clustering algorithms are probabilistic, since they are focused on model matching, using probabilities as opposed to distances to decide clusters (i.e., Expectation Maximization – EM- is an example of the probabilistic clustering algorithms).

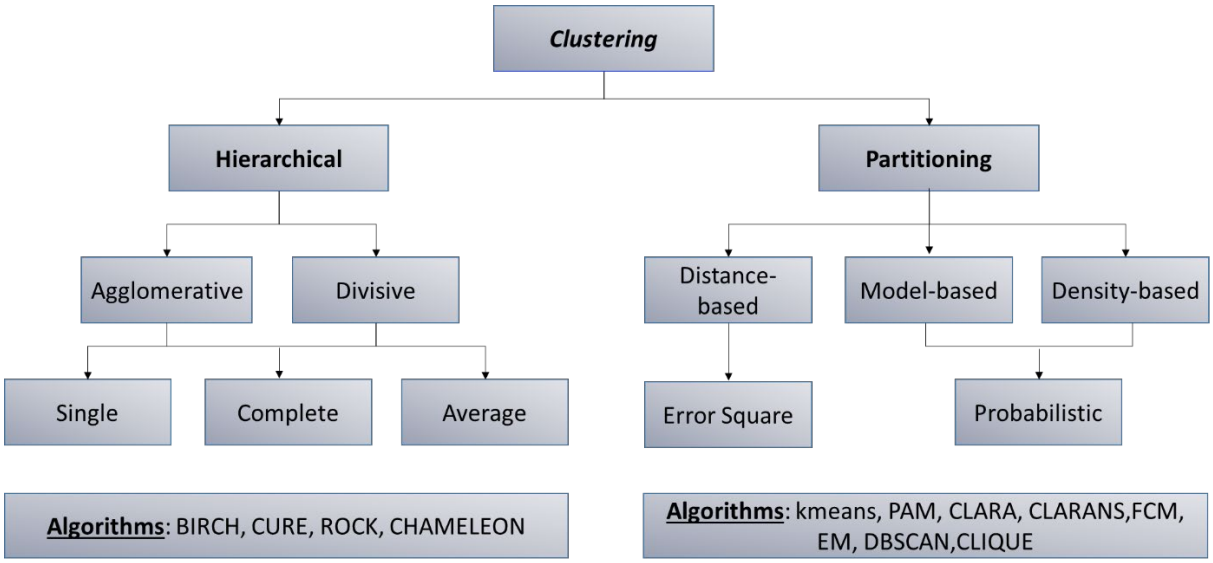


Figure 25. Categorization of Clustering Algorithms

Anomaly Detection

Recently, significant advances of **Anomaly Detection**[29] have been reported in diverse research fields, including Data Mining and Machine Learning. Anomaly Detection techniques aim to identify those regions from data whose behaviours or patterns do not conform to the expected values. These unexpected behaviours, which significantly differ from the normal ones, are called anomalies. An anomaly is also referred to as an outlier, a discordant object, or an exception, based on the specific application. Identifying abnormal or unexpected patterns is especially important to many technology domains, such as decision making, business intelligence, and data mining, among others. **Figure 31** illustrates a high-level overview of a typical anomaly detection process. The anomaly detection algorithm first receives the input data, then it builds a behavioural model from the training data and further applies the model to detect anomalies. If an anomaly is detected, the algorithm applies the similarity metric/ measure to identify the anomaly, in any other case, the algorithm does not produce any anomaly. Various supervised and unsupervised Anomaly Detection algorithms have been applied to enhance the various system's performance in all

[29] https //www researchgate net/profile/SalimaBenqdara/publication/325049804_Machine_Learning_Techniques_for_Anomaly_Detection_An_Overview/links/5af3569b4585157136c919d8/Machine-Learning-Techniques-for-Anomaly-Detection-An-Overview pdf

levels such as in clustering, features selection and classification.

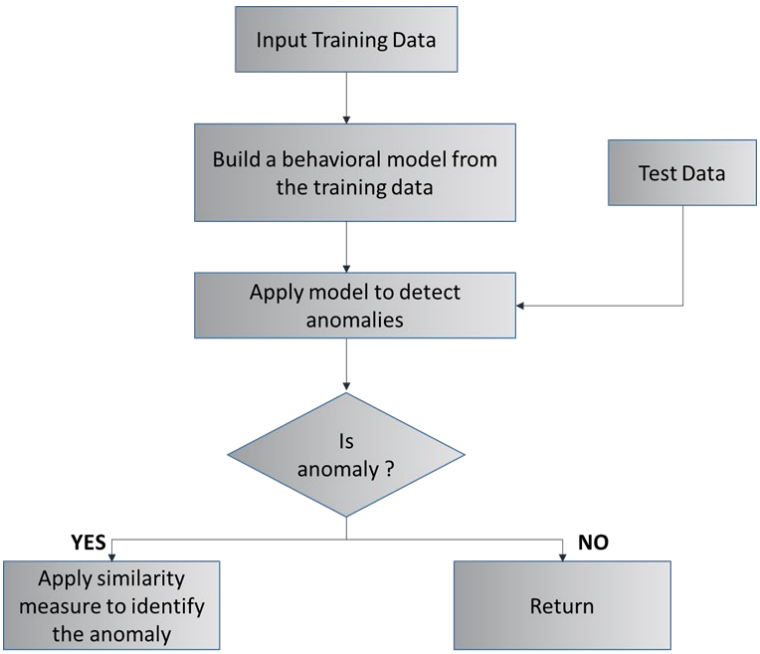


Figure 26. Anomaly Detection Process

The following table provides a thorough comparison among the different Anomaly Detection algorithms[30], with respect to their advantages, drawbacks, and the learning type. On this direction, **Anomaly Detection** contains both **supervised** and **unsupervised** learning techniques[31], while several challenging algorithms are proposed and used to achieve good results for these techniques. In the scenario where the test data contains no unknown anomaly types, supervised learning methods significantly outperform the unsupervised ones. However, based on the nature of the applications, different anomaly detection techniques may be used to detect effectively and with high accuracy all anomaly classes.

[30] https //www hindawi com/journals/complexity/2019/2686378/
[31] https //www researchgate net/profile/Gyoergy-Kovacs3/publication/326707960_A_taxonomy_and_platform_for_anomaly_detection/links/5c378fd1458515a4c71b7726/A-taxonomy-and-platform-for-anomaly-detection pdf

Technique	Learning Type	Pros	Cons
K -Nearest Neighbour	Supervised	• Easy to understand when there are few predictor variables. • Useful for building models that involve various data types, such as text.	• Large storage requirements. • Sensitive to the choice of the similarity function that is used to compare instances. • Lack a principled way to choose k, except through cross-validation or similar. • Computationally expensive technique.
Neural Networks	Can be either supervised (i.e., considering the scenario of labelled training data in the dataset), and unsupervised (i.e., where no labelled data exist in our datasets)	• Neural Networks can perform tasks that linear programming is not able to implement. • Learns from the input data • Can be implemented in any application • Robust	• Neural Networks need enough training observations to operate and produce high-quality results. • Usually require high processing time for large networks • Different architectures need to be investigated based on the nature of each application.
Decision Trees	Supervised	• Simple to understand and interpret. • Requires little data preparation. • Able to handle both numerical and categorical data. • Able to validate a model using statistical tests. • Robust • Perform well with large data in a short time.	• The problem of learning an optimal decision tree is known to be NP-complete under several aspects of optimality and even for simple concepts. • Decision-tree learners create over-complex trees that usually do not generalize properly over the input data.
Support Vector Machines (SVMs)	Supervised (can also be utilized in unsupervised learning problems)	• Find the optimal separation hyper plane. • Can deal with very high dimensional data • Present high-performance	• Require both positive and negative examples. • The performance is strongly related to the selection of a proper kernel function. • High memory and CPU requirements.
Self-Organizing Map	Unsupervised	• Simple and easy-to-understand algorithm • Topological clustering unsupervised algorithm that works with nonlinear data sets. • Excellent capability to visualize high-dimensional data onto 1- or 2-dimensional space, and thus it makes it a very good candidate for dimensionality reduction applications.	• Time consuming algorithm

Technique	Learning Type	Pros	Cons
K-means	Unsupervised	• Low complexity	• Necessity of specifying k • Sensitive to noise and outlier data points • Clusters are sensitive to initial assignment of centroids.
Fuzzy C-means	Unsupervised	• Allows a data point to be in multiple clusters.	• Need to properly define c, and the clusters number. • Need to determine membership cut-off value. • Clusters are sensitive to the initial assignment of centroids.
Expectation-Maximization Meta	Unsupervised	• Can easily change the model to adapt to a different distribution of data sets. • Parameters number does not increase with the increase of training data.	• Slow convergence in some cases

Table 18. Comparison of the most representative Anomaly Detection techniques

Since the current report is a study and does not involve at the current stage any implementation steps; we are not able to select a specific approach or algorithm for Anomaly Detection. When this study will be transformed into an implementation project, a complete benchmarking among the different techniques will be made and the proper algorithm for each scenario will be selected (according to their performance evaluation).**Conversational Agents and Virtual Assistants**

Conversational agents (i.e., **chatbots**)[32] and **virtual assistants**[33] are considered one of the key Natural Language Processing (NLP)-based players in the public sector[34]. Chatbots utilization in the public sector has the goal to improve service delivery and customer satisfaction, by simultaneously reducing the workload of the human-interaction and improving service delivery. Going online to access a public service, users are likely to encounter several pop-up assistants, which are not "real persons", but chatbots supported by **AI solutions**. Public agencies introduce chatbots for multiple reasons, including resources/ human-effort saving, and money saving, among others, since such technologies are designed to provide automated answers to frequent and simple queries.

In the current study, a chatbot-like solution (*BUC.06 - Virtual Assistance to Provide Guidance on Certain Activities*) will enable a natural-language search engine to find targeted answers to certain eu-LISA's, their stakeholders, and authorized user's questions. The main business value for such solution is to provide guidance and support on requests from several authorized actors, including travel applicants and/or risk profiles officers. Additionally, other actors may be added upon eu-LISA's business needs.

Nevertheless, due to the nature of how the chatbots are designed and operate, along with potential impacts on the public sector and specifically, eu-LISA, they are frequently confronted with social, ethical, legal, and political concerns. The key reason that lies behind this conclusion, is that users of such services, and the wider public, do

[32] https //www researchgate net/profile/Ebtesam-Almansor/publication/333931397_Survey_on_Intelligent_Chatbots_State-of-the-Art_and_Future_Research_Directions/links/5d47688a299bf1995b66374a/Survey-on-Intelligent-Chatbots-State-of-the-Art-and-Future-Research-Directions pdf

[33] https //link springer com/article/10 1007/s12525-020-00414-7

[34] Tambouris, Efthimios "Using chatbots and semantics to exploit public sector information " EGOV-CeDEM-ePart 2018 (2018) 125

not always feel that chatbot solutions demonstrate and adhere to certain public service values. However, the notion of "value" is much broader than simply saving money for the various organisations which aim to include such solutions. It corresponds to ideas such as trust, fairness, and transparency, among others. These values can be of great benefit to persons which are directly responsible for the development and delivery of public service chatbots, along with the users and the proper stakeholders that will exploit such solutions. Consequently, service values can be defined as a collection of social, professional, ethical, and other values that facilitate reasonable, legitimate, and relevant actions in the public sector, while they can be divided into several categories and dimensions (i.e., customised for eu-LISA's purposes), as illustrated on Table 19.

eu-LISA Related Service Value	Definition
Adaptability	*The degree to which the chatbot adapts itself to varying conditions (i.e. including changes in business rules, and varying technical conditions such as adapting to different devices and networks) when being used in delivering a service.*
User orientation	*The degree to which the chatbot accommodates user expressions and needs when being in use for resolving a case*
Professionalism	*The degree to which the chatbot displays a principled, competent, honest, respectful, consistent, and trustworthy conduct when it is being used in delivering a service.*
Effectiveness	*The degree to which the chatbot produces an expected outcome given the resources invested when it is being used in delivering a service.*
Efficiency	*The degree to which the chatbot facilitates service delivery while minimizing the cost and resources required.*
Fairness	*The degree to which discrimination (based on individual differences) does not exist when the chatbot is being used in delivering a service.*
Legitimacy	*The degree to which the chatbot complies/conforms with lawful and reasonable steps and mandates when being used in delivering a service.*
Acceptability	*The degree to which the chatbot is a viable option for service delivery and is beneficial in a way that public shows minimal or favourable reaction to using it.*
Openness	*The degree to which chatbots transparently disclose their identity to users before a service interaction starts and provide the rationale for decision-making when delivering a service.*
Accountability	*The degree to which the chatbot represents an accountable/responsible channel (including acknowledgement of limitations) when being in use for delivering a service.*
Social license	*The degree to which the chatbot acquires continued approval as a viable service delivery channel from the community and other stakeholders.*
Privacy	*The degree to which the chatbot ensures protection of user's information during and after being in use for delivering a service via any device.*
Collaborative intelligence	*The degree to which a chatbot partners with the user and other service stakeholders, complimenting their skills to fulfil service needs.*

Table 19. Service Values related to eu-LISA's Virtual Assistance Case

Chatbots and virtual assistants commonly rely on heavy Deep Neural Network (DNN) computations, Natural Language Processing (NLP) technologies, Generative Models (i.e., such as Generative Adversarial Networks – GAN's), and Reinforcement Learning (RL). However, the main limitation that organisations around the world confront, is that to develop effective and robust conversational agents' systems, they need to invest significant resources in order to provide high quality training data. Chatbot solutions require enormous amounts of training

data to provide accurate results. On this direction, multiple libraries (i.e. libraries of the decided programming language that will be used for any AI/ML implementation) and pre-trained NLP-based models are implemented to cover the demands. Another limitation of such technologies regards the learning language. Certain languages, such as English, have a significant corpus of textual information that can be used to train neural networks, whereas other minor languages, might have limited corpus of data that can be used for training, which may negatively affect the quality of the algorithms. However, as virtual assistant's technology evolves, the above-mentioned limitations are continuously resolved.

Considering the big picture, **conversational agents** and **virtual assistants compose** a **highly important solution for eu-LISA**, able to tackle several tasks **including IT service desk automation** and **providing support to the users of the various CBSs (i.e., ETIAS, VIS, SIS, EES, etc.)**. An in-scope, possible implementation of chatbot solutions in the current study, is described on "**BUC.06 - Virtual Assistance to Provide Guidance on Certain Activities"**, in which novel and flexible **Open-Source AI methodologies** and **tools** are proposed for the design of a **robust chatbot solution**.

5.2.3 Validation Step

The final step of the (**DM)ADV** Six Sigma framework, includes the validation process of the current analysis, including the list of the proposed AI tools per Business Use Case, which address all needs, business, and stakeholders' requirements as they were identified on **Phases 1** and **Phase 2** of the study. Additionally, all comments from the previous iterations, and during Phases 1 and 2 are tackled and considered in the final version of this deliverable. The validation process on *how AI and advanced analytics technologies can cover the demands of such a challenging project and organization*, is thoroughly described in **Section 5.3** and **Section 5.4**, which highlights the **assessment** of the **AI-enabled Business Use Cases**, and further explains Lot1 **suggestions** and **best fit solution**.

5.3 Data Science/ Machine Learning (DSML) Solution for eu-LISA

Nowadays, AI is considered a top priority for enterprise architecture and technology innovation, accompanied with several AI initiatives that continuously evolve across the various organisations. Nevertheless, moving such initiatives to production usually face significant challenges. For this purpose, **eu-LISA** should establish a strong and cohesive **AI engineering discipline** to benefit from **AI's governance**, **agility**, and **scalability capabilities**. On this direction, AI engineering methods are highly prominent to enable better governance and consistency in reusing, retraining, rebuilding, interpreting, and explaining the AI models, relying on the foundational pillars of **DataOps**, **DevOps**, and **MLOps**.

As part of the present study, we outline top engineering practices that accelerate the **operationalization of AI** [35] (**Figure 32**) to achieve business value for **eu-LISA's Business Use Cases**, as they are defined within **Chapter 3 - Business Use Cases** of this report. Specifically, we highlight the disciplines, objectives, and key players associated with the AI engineering perspective considered for **eu-LISA's** case, and in **CRRS**.

Following this direction, Business Analysts, AI architects and **eu-LISA's** stakeholders and/or the contractors, should stay involved in all aspects of the AI initiative, by gathering business goals and aligning the end objective to the strategic vision (i.e. of including AI technologies and tools in CRRS), and to the BUCs definition and refinements. Additionally, alignment with the various eu-LISA's engineering teams - in terms of defining success metrics, KPIs identification, cost-benefit analysis and performance monitoring of the AI initiative-, is highly crucial towards the proper functionality of the current study. Another critical pillar refers to the considered data engineering practice,

[35] Gartner, A CTO's Guide to Top Artificial Intelligence Engineering Practices Published 29 October 2021 - ID G00744628

which makes data from the various **CBSs** accessible and available to the various data consumers, such as ML engineers, data scientists, etc., by simultaneously informing all business stakeholders of the data situation at CRRS.

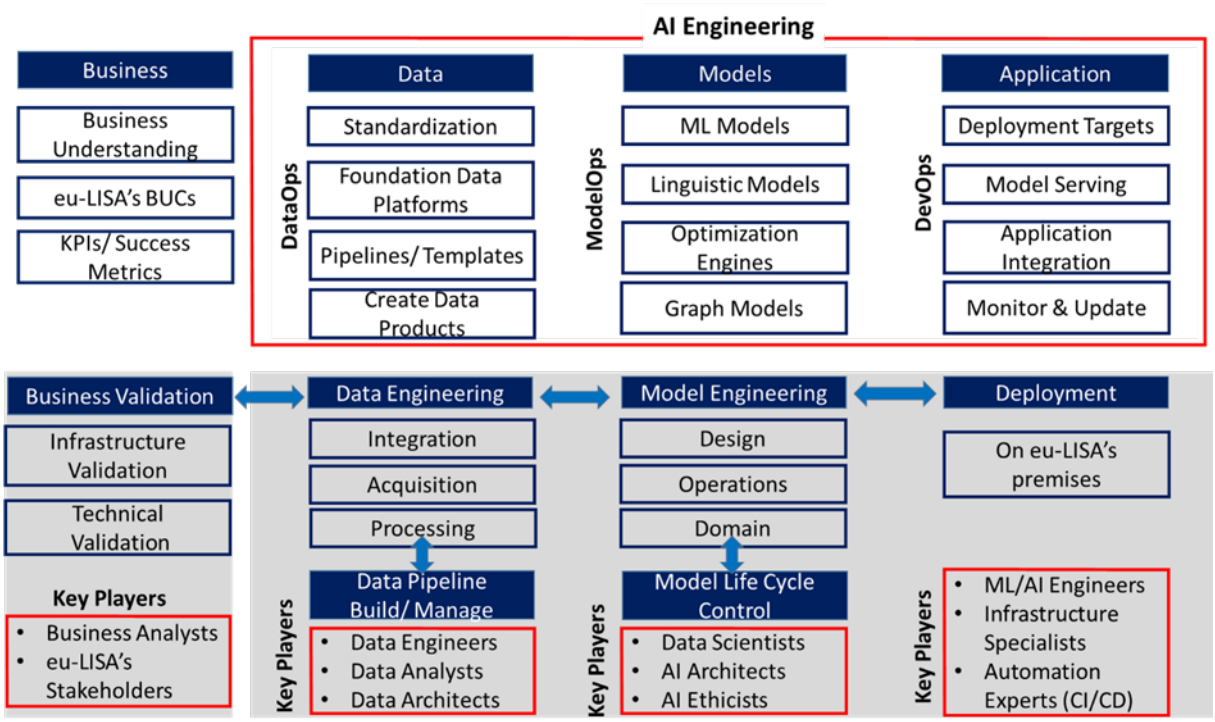


Figure 27. Operationalizing AI within the organization

Lot1 suggests the design of an efficient data engineering approach, containing the following steps:

- Adopt a Development Operations (**DevOps**) approach to data through automated build, test and/continuous deployment automation to manage the full life cycle of data pipelines.
- Treat data within **CRRS** as a corporate asset and ensure there is a centralized data management team (consisting of data engineers, database administrators, etc.,) to define standards for data management, preservation, and quality, and further adherence to eu-LISA's regulations.
- Empower line of business teams so they can manage their data and AI products by adapting Data Operations (**DataOps**) practices within the defined **BUCs**.

From the model engineering perspective, we suggest the generation of proper **AI/ML** models applying on the different **BUCs**. Model engineering is central to **AI engineering** since it converges various **AI platforms** and **solutions**, by simultaneously ensuring **reusability**, **scalability** and **governance** of the **AI models**, in a way that they interact with the business, data, and IT teams. Lot1 recommendations towards an efficient model engineering approach are summarized as follows:

- Ensure operationalization and scalable value generation, by aligning the model selection and training phase with the key business objective of the individual **BUCs**.
- Consider a centralized Model Operations (**ModelOps)** solution that encompasses continuous model deployment, validation checks, reusability, real-time monitoring, ongoing compliance and comprehensive change management capabilities.
- Evolve the **ModelOps** solution to make the model engineering practice even more future resilient, by performing quality checks at each step of the process, and by ensuring the outcomes explainability.

Finally, analysing the deployment (i.e., application engineering) perspective, we result in the conclusion that AI projects are by far different from software projects, due to the rapid iteration of models and experimentation with both model code and data. For this purpose, **DevOps** tools can assist in model service granularity, code versioning, and automated software delivery, consisting also a main part of the **AI development life cycle**.

Lot1 proposes an efficient deployment approach with its key steps to be summarized as follows:

Use **code repositories** that assist with source code management and container build/ management tools, which simplify model packaging, scanning, orchestration, and distribution.
Consider **Continuous Integration/Continuous Delivery (CI/CD)** tools that automate and accelerate software delivery. Model-serving tools streamline the deployment of AI models to a target environment, such as Kubernetes.
Ensure continuous delivery of **AI solutions**, creating small and self-organized teams, and by encouraging AI developers/ engineers and data scientists to closely work with **eu-LISA's stakeholders** and **Business Analysts**.
Consider an **end-to-end integrated DevOps toolchain** to bring standardization, governance, security and automated workflow into the AI development life cycle.

5.3.1 DSML Architectural Design

In the current analysis, we consider data from various data sources (i.e., **CBSs**), with different structures, and assume that both **batch** and **streaming data flows** will be investigated. For this purpose, different ingestion, integration, storage and analysis techniques are considered. As the current architectural landscape of **CRRS** is already explored, we put our effort in suggesting several **state-of-the-art**[36] and widely used Open-Source technologies in several fields of industry[37]. The following diagram highlights the main logical architectural view of our current analysis, and further illustrates how the various data sources can be combined through the **Data Ingestion**, **Data Storage**, **Metadata Management**, **Data Analysis**, **Artificial Intelligence & Machine Learning layers**, to be applicable for all **CRRS's Business Use Cases (BUCs)**. In the below figure, there are two areas highlighted in red. The first four processes (Data Sources, Data Ingestion, Data Store & Data Processing) belong to the current scope of CRRS, while the latest (AI/ML) is the one that was thoroughly studied in the current document.

[36] Gartner, Best Practices for Designing Your Data Lake Published 20 May 2021 - ID G00747336
[37] OpenDataHub OpenDataHub

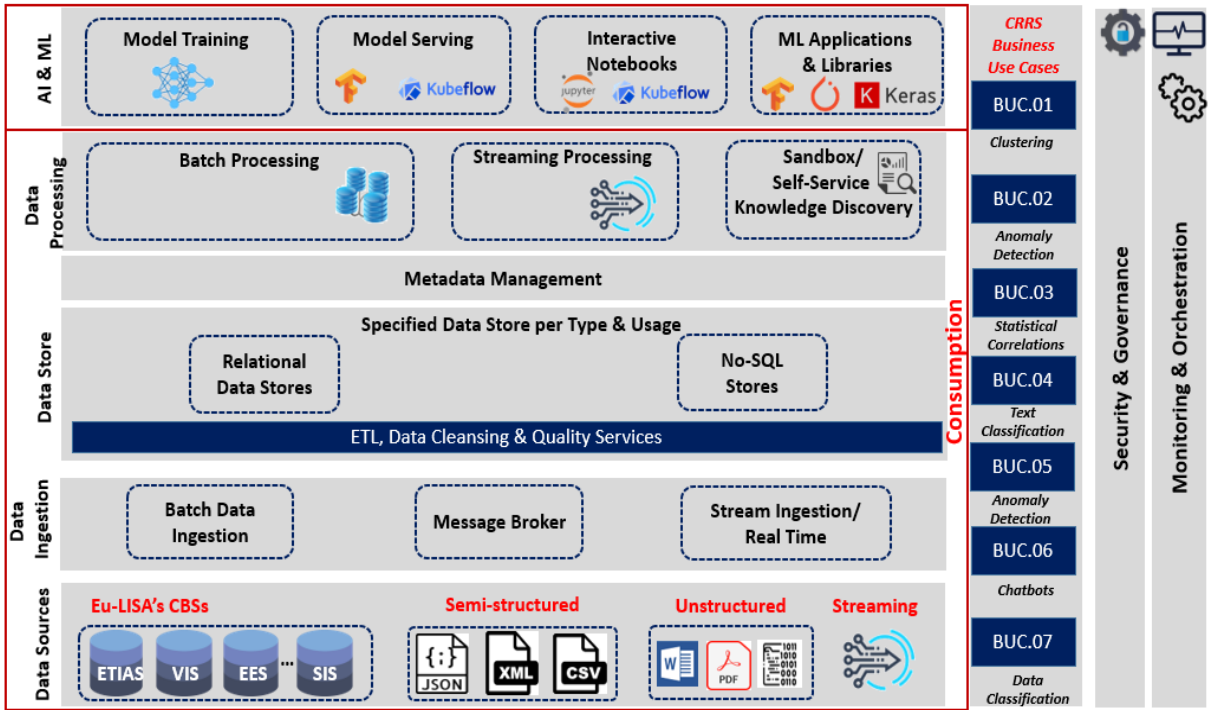


Figure 28. Data Science & Machine Learning (DSML) Logical Architectural View for CRRS

The components depicted in the above figure are explained further below:

- **Data Sources:** In the current analysis we consider data extracted from the various **Central Business Systems** (**CBSs**), and landed into the **CRRS**, such as: **ETIAS, VIS, EES**, and **SIS**, among others. Additionally, we consider data extracted via a separate/ **external data channel**, including reports uploaded by **authorized Stakeholders**, **Member States**, and **EU institutions**, among others. Characteristic example are the reports that reach CRRS from **WHO** and **ECDC**. The data sources are either **structured**, **semi-structured**, or/and **unstructured**, while they might reach **CRRS** in **batches** or in a **streaming** manner. Data in CRRS landing zone are passed through the corresponding CRRS's **Data Anonymization component**, which is triggered, i.e., deployed on **CBSs** sides. However, for the external data sources (i.e. reports which are uploaded into CRRS), the authorized users need to ensure that ***they will not reveal and further upload any sensitive information able to identify the individual's personal information***. Consequently, the proper identification of the data sources is the initial and highly significant step towards the design of the current architectural landscape since different types of databases and store systems will be suggested and further considered.
- **Data Ingestion & Integration:** The part where Data Ingestion has a prominent place is when combining data from the various data sources and systems and merge them. Additionally, at this step the required attributes that shall be used for analysis purposes are selected. After the ingestion part is completed, the next step is an ETL (Extract - Transfer- Load) process, which will include pipelines responsible for transferring data to the appropriate storage database(s) as well as handle any data transformations (e.g., handle missing values, filtering of data, create new columns, and data aggregation, among others).
- **Data Store:** One of the main base components in every AI/ ML-related project is data storage. For serving data coming from various analysis and in different formats, it is essential to suggest a data storage able to support this variety/ assortment. For instance, when running text processing tasks and conclude to results in txt/xml/csv form, the data storage type should have considered this case even if it was not initially designed for such data types. Specifically, in the current study we consider two main different data store categories: (i) **relational data stores**, and (ii) **NoSQL data stores**. Both categories include various tools that support them, and they are described thoroughly in our subsequent analysis. According to the already existing Software Architectural landscape of CRRS, the persistent storage solution which are

drawn from the CSI Technology Portfolio Catalogue include: Kafka, CEPH, PostgreSQL - with the addition of: (i) a shared file system for storing generated reports and uploaded documents, data quality and extraction rules per CBS; (ii) IBM Data Virtualisation to provide a uniform, "fast-query-able" abstraction of the data in CEPH S3 and PostgreSQL, which is required for running SQL analytics on all sources of data used for reporting.[38] However, in the current analysis, we design solutions dedicated to the Artificial Intelligence / Machine Learning (AI/ML) layer, relying only on Open-Source frameworks. It is crucial to highlight that the data store technologies are complementary and further enable the proposed AI solution. Although the current study is focused on AI technologies, it is crucial to propose these complementary tools (even if they are out of the scope of the current study). An in-depth analysis of the complementary technological frameworks and tools in provided on Appendix I - Proposed Technological Frameworks, which considers tools for all different building blocks of the DSML platform. To summarize. Lot 1 suggestions for the persistent Data Store layer arise from our insight to provide also with flexible and non-obsolete solutions.

- **Metadata Management**: For designing and proposing a solid yet reliable solution, metadata management operates as a key principle for an organized structured that will also consider any future needs of the organization. Metadata information shall be arranged and ordered properly to be accessible to specified users with grand access rights. However, as part of this study metadata management is formed in a slightly different way, since it is focused on mainly data quality tasks and how they should be handled.
- **Data Processing:** This layer includes processes such as collecting and manipulation of data in each step of the procedure. It is considered as the orchestrator, which ensures the proper operation of the system. Specifically, it includes tools for both streaming and batch data types, while it also it provides data exploration capabilities, within the corresponding Sandbox, i.e., self-service knowledge discovery tools.
- **Artificial Intelligence & Machine Learning:** The concepts of AI and ML are the foundation layer of the current analysis for CRRS. In this layer, we consider all technological components and tools that are required to encapsulate Lot1 proposed **DSML** architecture. Model training tools and libraries are the first pillar in suggesting a proper AI & ML architecture, while model serving tools will be responsible for providing immediate inference for all BUCs, for eu-LISA's key stakeholders to perform accurate and prompt decision making. On this direction, interactive notebooks will host the development and further deployment of the ML/AI and data science algorithms for the various ML-based applications. Finally, Business Intelligence provides the main interaction between the Business Analysis and AI part.

The above-mentioned layers are the key pillars towards the proposed solution. However, to provide a comprehensive approach, we provide a brief overview of two extra layers that may be considered in the near future but are not part of the current study, namely: the Security & Governance layer, and the Monitoring & Orchestration.

- **Security & Governance:** In this layer the main security and governance practices are considered. Data in all situations, either in storage, or in motion, require adhering to concrete security rules. However, security and governance are not key concepts of the current assessment, and for this purpose we stay in line with the already existing technologies and tools for Security and Governance used by the organization (eu-LISA). To provide a highly flexible and scalable Open-Source AI architectural Solution that will be able to securely host the various BUCs, Lot1 suggests CRRS DSML platform to consider Ceph Object Gateway[39] tool, providing encryption of uploaded objects and options for the management of encryption keys. Additionally, Keycloak[40] is a candidate, providing user authentication for eu-LISA's personnel and key stakeholders.
  Monitoring & Orchestration: This layer is responsible for monitoring all aspects of the proposed end-to-end AI platform. Since monitoring & orchestration of the DSML platform for CRRS is not the main scope of the current

[38] CRRS-E2-WP2-001-1 - Software Architecture Document (SAD) / Architecture Overview

[39] https //docs ceph com/en/latest/radosgw/s3/
[40] https //www keycloak org/

study, we stay in line with the current Architectural Landscape of CRRS, and specifically eu-LISA's CSI catalogue. Specifically, for Deployment and Orchestration considered tools are Ansible Tower[41], Terraform, JFrog Artifactory. For Monitoring and Observability, Prometheus[42], Splunk, EFK are utilized, while for building and executing post-build steps (i.e., tests) Jenkins[43] is considered, providing the functionality to create and manage workflows for build and release automation.

To further support the selection of the appropriate tools for the various layers we perform a thorough comparison for the various building blocks. For this purpose, Appendix I - Proposed Technological Frameworks is dedicated to described and perform an in-depth analysis of the most dominant Open-Source tools/ technological frameworks, per building block considered for all BUCs. Last, in the individual section of each the BUC, we explain in further detail our best fit solution.

5.4 Assessment of the AI-enabled Business Use Cases in eu-LISA

Key objective of the current study is the identification of the proper **AI Solutions** (i.e., **AI tools** and **algorithmic formulations**) which are prominent to enable the identified **Business Use Cases** (**BUCs**) within the proposed **Data Science & Machine Learning (DSML)** platform. On this direction, it is deemed necessary that Lot1 provides an overview of the key specifications, able to host the described **AI frameworks** and **tools**. To support the surrounding building blocks and to stay persistent with the already existing architectural landscape of CRRS, and further avoid hardware constraints, we propose to support our analysis on the architectural principles of **containerization** and **microservices, as well as the main principles of eu-LISA**.

Containerization shall enable **eu-LISA's** authorized users to execute the **AI-enabled BUCs** under isolated spaces referred as containers. While virtualization allows running multiple discrete Operation Systems (OS) on the hardware, containerization enables the deployment of several applications using abstraction on an OS layer, while it simultaneously provides high levels of elasticity and scalability.

On this direction, a container platform is a form of virtualization technology that significantly facilitates the development and deployment of several applications inside packaged virtual containerized environments. Containers that can be deployed to any machine without depicting compatibility issues, may be considered. In the case of a container platform, a Linux/Unix-based virtual machine can also be installed on the host machine. In contrast with virtual machines, containers do not require memory or disk space pre-allocations since the containers exploits the already existing resources. Additionally, microservices is an architectural approach to application development in which a single application is composed of many small and autonomous services. Microservices are independently deployable, enabling **eu-LISA's** users to improve their developments, add new features/ updates, and add extra levels of scalability compared to a monolithic architecture.

When working with a sizable number of containers, it is important to use orchestration tools that automate operational tasks, such as distributing containers across a cluster of servers. On this direction, Lot1 investigates the exploitation of OKD[44] and Docker Containers[45] with Kubernetes[46]. Specifically, Kubernetes is typically the go-to container orchestrator, providing great level of extendibility and customization. It is a portable, extensible, open-source platform for managing containerized workloads and services, providing resilience, with scaling and fail-over for the various applications. Kubernetes ensures that each one of the containers is where it is supposed to be and

[41] https //www ansible com/products/controller
[42] https //prometheus io/
[43] https //www jenkins io/
[44] https //www okd io/
[45] https //www docker com/
[46] https //kubernetes io/

that they are working together properly. Additionally, it is crucial to highlight that in the current study Lot1 investigates the exploitation of the most appropriate and widely exploited open-source tools and frameworks, able to support the proposed DSML solution. In subsequent assessments, the suggested tools/ frameworks are expected to be integrated with CRRS current architectural landscape and the already existing components. Consequently, the precise determination of the components that will be used on the solution, regards a future design assessment, which should be implemented in close cooperation with eu-LISA's personnel/ Solution Architects/ System Architects, etc.

The **DSML** solution also considers a persistent **Data Store layer**, in order the respective frameworks and tools to interact accordingly and retrieve or store necessary data. The **Data Store layer** shall store the different data types, i.e., **hot** (i.e., **near-real time/ streaming data**) and **cold path** (i.e., **batch/historical data**), as they are extracted from the various **CBSs**. It is deemed necessary that this layer should persistently store the results from the model training phase (i.e., **models/ models' metadata information**), along with the results from the model inference phase (i.e., **predictions from the model testing phase**, when new data reach the DSML platform, applying for each BUC). On the top of the architecture, Lot1 suggests the design of three individual, isolated environments: **Development** (**Dev**), **Testing** (**Unit/ component level**), **and Production** (**Prod**). In the Development environment the main AI code development tools will be hosted, and it is further the responsible environment for performing all necessary **ML/AI training** processes steps applying on the various BUCs. Additionally, the evaluation metrics (i.e., training/ validation accuracy, absolute reconstruction errors, etc.) will be extracted for the AI engineers/ data scientists to test the different algorithmic formulations and select the best possible algorithm fitting to the BUCs. Testing environment is responsible for all required unit and component testing of the various AI solutions within the DSML platform. Finally, Production environment is responsible for deploying the ML/AI algorithms (i.e., model serving), and extracting the model inferences tested with new data that reach the DSML platform.

The following paragraphs provide a thorough analysis on the **AI Solutions** proposed for the **individual BUCs**, along with Lot1 **best fit recommendations satisfying each of the BUCs.** The recommended AI solution *complies with our existing knowledge on eu-LISA's computing environments, and specifically aligns with the CRRS, data architectural landscape*.

The presentation of each **Business Use Case** and the recommended **AI solution** follow a specific format. First, there is a small description of the selected **Business Use Case** along with the **datasets/ data structure** that will be used during future implementations. Furthermore, the suggested **AI methodology for each BUC,** is explained (i.e., regarding the selection of the proper AI/ML algorithms) and the indicative **implementation steps** are outlined. Lastly, to propose robust **AI Solutions**, it is essential to discuss about the **Key Success Factors/ Pre-requisites** required for this analysis, along with the **assumptions** and **constraints**.

Finally, it is important to note that according to the Stakeholder Requirement "REQ-50.01.02 (SHR): The CRRS shall comply with Accountability Principles for Artificial Intelligence", the proposed **AI Solutions** corresponding to the individual BUCSs, will comply with the ***Accountability Principles for Artificial Intelligence (AP4AI) in the internal security domain*** , including the following aspects: Legality, Universality, Pluralism, Transparency, Independence, Commitment to robust evidence, Enforceability and redress, Compellability, Explainability, Constructiveness, Conduct, and Learning organisation.

5.4.1 BUC 01 - Identification of Risk for a Specific Group of Travellers

BUC-01

Name	IDENTIFICATION OF RISK FOR A SPECIFIC GROUP OF TRAVELLERS
Description	The purpose of this business use case is to enhance the **CRRS** analytical capabilities beyond the deterministic approach, by using AI technology to **support human decision making**, as it was initially described on Section "BUC 1: AI driven identification of risk for a specific group of travellers." During the definition of the risk indicators, the proposed AI technology will **identify patterns** or a **set of common characteristics** from the analysis of **historical data** available in the **CRRS** (collected from the CBSs) pointing to **security**, **illegal immigration** or **high epidemic risk** for a specific group of travellers (**Clusters**). For instance, different groups of travellers might indicate epidemic risks (i.e. low, medium, high). However such scenarios should be commonly agreed with the stakeholders and should be also included on the corresponding requirements section. NOTE: in this use case, the requirements of preventing direct or indirect discrimination in the application of the data sets are considered; and ensuring that specific risk indicators must be targeted and proportionate. See Section 9.1.1.4 and Section 9.8.1.1
Data Structure/ Dataset	**Structured** and **Anonymized data**, including attributes extracted from the various CBSs (i.e., VIS, EES, ETIAS[47]) such as: { Status Information", "VISA Authority", "Location of Visa Authority", "Nationality", "Member State", "Date of VISA Application", "Place of VISA Application", "Type of VISA", "Type of Travel Document", "Number of VISA Applications" , "Status of VISA Application", "Date of Acceptance", "Date of Refusal", "Purpose of Journey", "Fingerprints", "Sex", "Year of Birth", "Date of crossing", "Border crossing point", "Three-letter code of the issuing country", "Number of Persons identified as over-stayers", "Nationalities of persons identified as over-stayers", "Revoked Stay", "Validity Extension", "Refused Entry", "Refused Entry Nationality", "Refused Entry Type of border", "Refused Entry border crossing point", "Refusal Reason", "Country of Residence", "Limited Territorial Validity", "Validity Period for travel authorization"}.
AI Methodology	**Classification approaches** will perform the identification of the groups of travellers corresponding to specific groups of risk. Classification categories can be defined based on specific indicators, i.e., based on similar age/education/occupation, etc. or any combination of the input attributes. The exploited **AI methodology** will rely on **Semi-supervised** and **Unsupervised** algorithms, based on the nature of the input data (i.e., whether labelled information exists). All Machine Learning (ML)-based approaches can use vector space models, where the concept of an 'embedding' (i.e., the numerical representation for how an AI intercepts data) is used to **partition data into groups of travellers (i.e., clusters)**. The consideration of the above-mentioned algorithmic categories is made to cover more clustering cases; however, upon the implementation of the specific BUC, and according to the provided data, precise determination of the algorithms is expected to be implemented. Characteristic examples of clustering algorithms are the following: Support Vector Machines Classifiers (SVMs)/ Decision Trees/ Random Forests/ Hierarchical Clustering, Deep Neural Networks are candidates for the specific BUC.

[47] SIS can be also investigated in near-future studies

BUC-01

Indicative Implementation Steps	**Indicative key steps of the implementation process are described below:** Merge the information from the different CBSs (i.e., VIS, EES, ETIAS). • Extract meta-data from the input data sources (i.e., dataset enrichment/ feature extraction/ dimensionality reduction, etc.). • Setup different scenarios, regarding the required risk-levels (e.g., very-high risk, high risk, medium risk, low risk) and further set risk threshold (in discussion with eu-LISA's experts & key stakeholders). Additional risk-levels related to security-risk, illegal immigration-risk or high epidemic-risk can be derived, upon common decision with the stakeholders. • The different scenarios will establish the final view of the grouping of travellers into groups of risks (clusters). • Split the dataset into training & validation sets using state-of-the-art practices. • Perform model training with different AI/ Machine Learning (ML)-based algorithms to select the best possible algorithm, via a thorough evaluation approach. • For each AI/ML-based trained algorithm, extract the corresponding evaluation metrics (i.e., model accuracy) to further validate the robustness of the proposed AI Methodology for BUC.01. • The algorithm which ranks the highest evaluation metrics and accuracy, for both training and validation sets, will be selected for BUC.01. **Process Output** • The output of this analysis is expected to **extract the groups of travellers** (i.e., in technical terms – **clusters**), along with the **assigned risk level**.
Key Success Factors/ Pre-requisites	**Key Success Factors/ Pre-requisites for BUC.01 are considered the following:** • Determine the appropriate number of groups of travellers (i.e., clusters), by either enforcing pre-defined number of clusters, or via a precise analysis which indicates them (i.e., according to stakeholder requirement: REQ-01.01.03 (SHR): The CRRS shall allow to determine the appropriate number of groups of travellers (i.e., clusters). • Definition of the risk-level severity. From a scientific point of view, this BUC relies on state-of-the-art clustering algorithms. According to standard algorithmic formulations and research on this topic, the optimal number of groups, extracted under a clustering process, may range from one to approximately six or seven groups, for achieving accurate results. Precise cluster characteristics (i.e., which information is required to be displayed within the extracted groups of travellers) – according to stakeholder requirement: REQ-01.01.02 (SHR): The CRRS shall allow to define the risk-level severity, upon agreement among eu-LISA's stakeholders. • Reflect the corresponding requirement for preventing direct or indirect discrimination in the application of the datasets, i.e. REQ-50.01 (BR): The CRRS shall not receive or process any personal data that could allow discrimination against persons on any grounds or violating human dignity, integrity and fundamental rights.
Assumptions	**Key assumptions for BUC.012 may be considered the following:** • All data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved.

5.4.1.1 Description of the Best Fit Solution

The purpose of this first Business Use Case is to support the human decision making by extracting groups of travellers (i.e., clusters) that have similar characteristic and behaviour assigned with the respective **risk level**. According to the stakeholders' requirements for BUC.01, a **robust AI solution** is designed in order to have access in related ETIAS data through CRRS, without allowing identification of any individual.

The related data attributes considered are: the application status information, the nationalities, sex, year of birth of the applicant, country of residence, education level, current occupation, type of travel document and three-letter code of the issuing country, type of travel authorization, validity period of travel authorization, and grounds for refusing, revoking, or annulling a travel authorization.

Specifically, in our design, we use **CRRS historical data** (incoming data collected from the different **CBSs**) and leverage them to identify the distinct groups of travellers. Via this approach, we derive an impartial clustering amongst the groups of travellers that will **fully respect all Fundamental Rights**. The risk indicators are not fixed yet and shall be strictly determined either by thorough discussions with **eu-LISA** or by utilizing previous experience gained from **eu-LISA's experts**.

This scenario is considered a typical **Machine Learning (ML)** classification use case where the result determines groups of travellers assigned with their risk level. Indicative list of severities can be refined or adjusted in coordination with the Stakeholder's needs. Additionally, from a scientific point of view, this BUC relies on state-of-the-art clustering algorithms; and according to standard algorithmic formulations, the association of the risk levels depends on the optimal number of groups that can be extracted under a clustering process, and after a thorough discussion with eu-LISA's stakeholders. Consequently, the precise clusters levels will be determined by the **resulting number of clusters**. The optimal number of clusters can either be pre-defined or use methods such as the **Elbow method** [48]or the **Silhouette method**[49]. The Elbow method illustrates in the form of diagram the explained variation as a function of the number of clusters and the optimal number is at the elbow of the curve; while the Silhouette method computes silhouette coefficients of each point to compare any similarity between the cluster that this point belongs and the rest of them. After the number of the clusters has been defined, the next step is to assign the definition of the risk-levels. As we mentioned earlier, the definition is closely dependent on the number of clusters however it will be discussed with **eu-LISA** team after the analysis has been completed. Consequently, **human intervention** is crucial, and mandatory for the successful use of the specific BUC.

In more details, **Figure 34** depicts the high-level process flow diagram for **BUC.01** and highlights the flow of the procedure and the role of each component. As a first step, historical data is inserted from the various CRRS sources, and is further used to initiate the analysis. Following the intermediate step, where data will be processed, cleansed, and any required feature extraction process will take place. At this stage, data will be analysed and prepared to be inserted in a **NoSQL** data store, as it is described in excessive detail on Appendix I - Proposed Technological Frameworks for both relational and NoSQL data stores.

After careful consideration regarding the **AI needs** of **eu-LISA** and the types of models that will suggested as part of this study, NoSQL data stores were selected. NoSQL data stores excel when it comes to elasticity and handling unstructured or semi-structured data. Thus, the data stores that were selected for the BUC.01 purposes were either **Apache Cassandra database**, or **MongoDB**, or **Redis**. When the data is stored in the respective data store, the respective algorithms are executed. To compare the different algorithms, the results of each one of them will be evaluated using various evaluation metrics. The clustering algorithms that were selected and considered as the most suitable ones, are the: **Kmeans**[50]**, DBSCAN**[51] and **EM**[52] (Expectation Maximization). All three of them belong to different categories, i.e., Kmeans is partitioning algorithm, EM is Graph-based and DBSCAN is density-based. More details regarding the selected algorithms can be found on section 5.2.2.1. The algorithms will be evaluated based on performance metrics such as Silhouette Coefficient, Dunn's Index etc. When the analysis is completed,

[48]https //www researchgate net/profile/HestryHumaira/publication/339670247_Determining_The_Appropiate_Cluster_Number_Using_Elbow_Method_for_K-Means_Algorithm/links/6142ce3a7d081355ccef105b/Determining-The-Appropiate-Cluster-Number-Using-Elbow-Method-for-K-Means-Algorithm.pdf

[49] https //www sciencedirect com/science/article/pii/0377042787901257

[50] Kodinariya, Trupti M , and Prashant R Makwana "Review on determining number of Cluster in K-Means Clustering " International Journal 1 6 (2013) 90-95

[51] Khan, Kamran, et al "DBSCAN Past, present and future " The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014) IEEE, 2014

[52] Ordonez, Carlos, and Edward Omiecinski "FREM fast and robust EM clustering for large data sets " Proceedings of the eleventh international conference on Information and knowledge management 2002

and the clusters have been configured, the model will be served using frameworks such as the ones described on Section 7.5 Open-Source AI/ML Tools. However, the selected frameworks as suitable for the current BUC are **TensorFlow serving**, **Torch Serve** or **KFServe.** Last, several visualisation tools have been considered to present the results of the clustering in a human-friendly manner. For this purpose, the most predominant candidates are **Pentaho** and **SuperSet**.

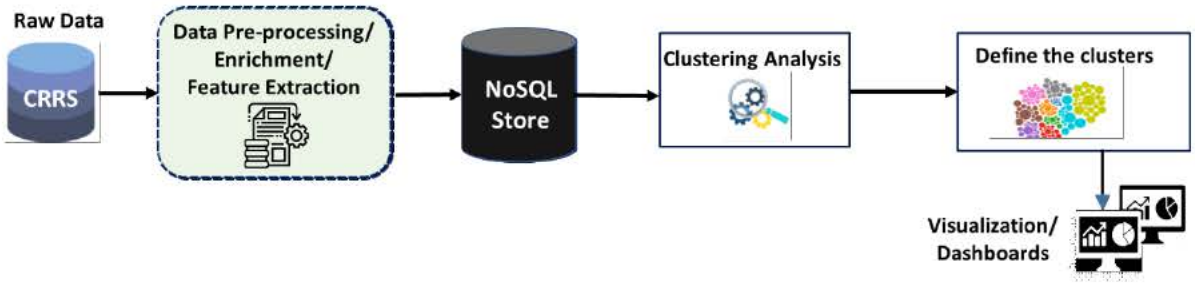


Figure 29. Proposed High-Level Flow Diagram for BUC.01

5.4.2 BUC 02 - Review and Validate Identified Groups of Travellers of Risk

BUC-02

Name	REVIEW AND VALIDATE IDENTIFIED GROUPS OF TRAVELLERS OF RISK
Description	The purpose of this business use case is to enhance the **ETIAS** analytical capabilities beyond the deterministic approach, by using AI technology to **support human decision making**. The AI technology will enhance the existing ex-ante assessment process for the already defined screening rules.
Data Structure/ Dataset	Structured and Anonymized data, including attributes extracted from **all CBSs** including (among others): { Status Information", "VISA Authority", "Location of Visa Authority", "Nationality", "Member State", "Date of VISA Application", "Place of VISA Application", "Type of VISA", "Type of Travel Document", "Number of VISA Applications" , "Status of VISA Application", "Date of Acceptance", "Date of Refusal", "Purpose of Journey", "Fingerprints", "Sex", "Year of Birth", "Date of crossing", "Border crossing point", "Three-letter code of the issuing country", "Number of Persons identified as over-stayers", "Nationalities of persons identified as over-stayers", "Revoked Stay", "Validity Extension", "Refused Entry", "Refused Entry Nationality", "Refused Entry Type of border", "Refused Entry border crossing point", "Refusal Reason", "Country of Residence", "Limited Territorial Validity", "Validity Period for travel authorization"}.
AI Methodology	In this Use Case we design an **AI Solution** that seeks for deviations from the already identified groups of travellers of risk for ETIAS. Statistical techniques will be considered, along with state-of-the-art time-series analysis methods. Specifically, we tackle the specific BUC, as a characteristic scenario of **Machine Learning (ML)-based Anomaly Detection**. Normal patterns of ETIAS risk profiles will be constructed and will further compose our solution's training dataset. Indicators illustrating the "**normal**" behaviour of the risk profiles will be commonly agreed with **eu-LISA's stakeholders**, and every actor (i.e., **risk profile**) that variates from this "**proper/normal**" behaviour will be marked as "**suspicious**". *Suspicious behaviour will be flagged and communicated to the responsible authorities (i.e., eu-LISA authorized users).*

BUC-02

Indicative Implementation Steps	**Indicative key steps of the implementation process are described below:** • Extract meta-data information from the various CBSs databases (i.e., dataset enrichment/ feature extraction, etc.). • Agree with eu-LISA's stakeholders on the particular risk profiles that will be considered and labelled "**normal**". • Agree with eu-LISA's stakeholders on the particular risk profiles that will be considered and labelled as "**suspicious**"[53] **(i.e. non-normal/ anomalous).** • To Generate the **Machine Learning (ML)**-based models, we split the dataset into **training** & **validation** sets using state-of-the-art practices. • Perform model training on the labelled "**normal**" and "**suspicious**" risk profiles using different Machine Learning (ML)-based algorithms and build the representative Machine Learning (ML) models. • Extract the corresponding evaluation metrics (i.e., **model accuracy**) for both training & validation sets to further validate the robustness of the proposed AI algorithmic formulations for **BUC.02**. • The algorithm which ranks the **highest evaluation metrics indicators**, for both the training and validation sets, will be selected for **BUC.02**. • For the testing phase, i.e., validation, when new groups of travellers reach ETIAS system, an inference will be assigned based on the trained model, indicating whether the specific group of travellers is predicted as "**normal**" or "**suspicious**". **Process Output:** This analysis will produce, for further utilization to the ETIAS group, an indicator reflecting a "**normal**" or "**suspicious**" behaviour/outcome, for a particular incoming group of travellers.
Key Success Factors/ Pre-requisites	Key Success Factors/ Pre-requisites for BUC.02 are considered the following: • Agreement of the anomaly detection levels (i.e., "normal", suspicious", or more levels, commonly agreed with eu-LISA's stakeholders) corresponding to the groups of travellers. (The reader may refer to stakeholder requirement: REQ-02.01.02 (SHR): The CRRS shall allow to define anomaly detection levels corresponding to the groups of travellers.) • Enough historical data to perform the analysis. (According to stakeholder requirement: REQ-02.01.01 (SHR): The CRRS shall ensure enough historical data to perform review and validation of identified groups of travellers of risk.) • Reflect the corresponding requirement for preventing direct or indirect discrimination in the application of the datasets, i.e. REQ-50.01 (BR): The CRRS shall not receive or process any personal data that could allow discrimination against persons on any grounds or violating human dignity, integrity and fundamental rights.
Assumptions	**Key assumptions for BUC.02 may be considered the following:** • All data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved.

[53] For further details refre to Section 9 2 2 3

5.4.2.1 Description of the Best Fit Solution

The goal of **BUC.02** is to validate whether the **identified groups of travellers of risk** for **ETIAS** are considered "**normal**", according to the specifications of **ETIAS** regarding the **validity of risk profiles**. According to the stakeholders' requirements formulated for the specific BUC, **Lot1** proposes the determination of the validity of risk profiles, without including any personal information. For this purpose, key prerequisite, as strongly implied on section 3.2.1 (i.e., Required Information – BUC.02), is to consider ***records from the CBSs indicating abnormal rates of refusals of travel authorizations***. As a result, the specific BUC not only relies on a robust and well-oriented **AI Solution**, but also incorporates human intervention from **eu-LISA's** and **ETIAS stakeholders** on the labelling of the input information. Specifically, one of the initial designs and (subsequently) implementation steps of the current analysis for **BUC.02** corresponds to the creation of a representative training dataset, including historical (i.e., batch data) risk profiles that can be considered and labelled "**normal**", and risk profiles which presents deviations from the expected behaviour or structure, and thus are considered and labelled as "**suspicious**" **(i.e. non-normal/ anomalous)**.

Upon common agreement with **ETIAS/ eu-LISA's** users on the characterization and further labelling of the historical risk profiles, we are able to tackle **BUC.02**, as a representative scenario of **Machine Learning (ML)**-based **Anomaly Detection**, adhering to semi-supervised principles, in the sense that partial knowledge regarding the anomaly type (i.e., "**normal**" and "**suspicious**"), appears on the training dataset. On this direction, patterns of risk profiles which are considered "**normal**" will be constructed and will further compose our proposed **AI solution's training dataset**. Consequently, when a new data instance reaches the system, the already trained model will be combined with the input data to provide the **prediction** (i.e., inference of the AI process). In case where the risk profiles variates from the learnt "**proper/normal**" behaviour will be marked as "**suspicious**". Suspicious behaviour will be **flagged** and **monitored** via the **responsible authorities** (i.e., eu-LISA authorized users).

From a technical point of view, **Figure 35** illustrates the proposed flow diagram for **BUC.02**. As we may observe, the specific BUC handles both batch/ historical data, extracted from the various **CBSs**, and streaming data that will be used for the *online anomaly detection process*. Input data observations, corresponding to the groups of travellers of risk, will be first pre-processed, to extract significant metadata information and representative features, necessary for the **AI model training phase**. Then, the processed data are ready for the training phase, and they are first stored on a **NoSQL** database, which can interact with the model training component. Specifically, the model training component is responsible for extracting the **AI models**, along with their evaluation metrics, indicating the **accuracy** with respect to the training dataset, other evaluation indicators, and the **robustness** of the **training phase**. The trained model will be served, via a **model serving** approach to the **online anomaly detection component,** i.e., **model inference phase**.

Every new data instance (i.e., risk profile from the CBSs) that reaches the system in **near-real time** conditions, will be combined with the trained model from the **offline model training** process, and **predictions** (i.e., probabilities) corresponding to either the "**normal**", or "**suspicious**" class will be assigned. The highest probability results into the extracted inference for the investigated risk profile. In the scenario where the risk profile is determined as "**suspicious**", Lot1 suggests being displayed on an interactive dashboard, able to be monitored by the responsible authorities (i.e., ETIAS, eu-LISA's authorized users). Additionally, in all cases, the results from the current analysis, along with the trained models, evaluation metrics, and models' metadata information will be also written into the **NoSQL** database.

Considering the enormous increase of data, and the probability of having low query-time from **eu-LISA's** systems, and specifically **ETIAS**, Lot1 suggests for each new data instance that reaches ETIAS, the exploitation of **Apache Spark Streaming** as the most appropriate tool for **(near) real-time processing**.

Spark Streaming provides its own high-level abreaction for continuous data streams, the so-called DStream, which processes data into mini batches, relied on a configured streaming interval. Additionally, in the scenario where the current architecture requires adjustments, Apache Spark module can analyse data on traditional databases with high efficiency, as well. Equivalent tools that can be used instead of **Spark**, are **Apache Flink** and **Apache Beam**, which also process both batch and real time data. For the data store part, **Lot1** suggests the selection of **NoSQL data store**, able to store the following data types: (i) processed data ready for the model training phase (e.g., features, extra columns, etc.); (ii) AI/ML models, along with their metadata information, (iii) and model inferences (i.e., predictions for each evaluated instance). On this direction, several remarkable candidates have been reported and thoroughly analysed on Section 0. For the specific **BUC.02**, **Lot1** suggests the adoption of either **Apache Cassandra**, or **MongoDB**, or **Redis**.

For the model training phase, several **AI frameworks/ tools** may be exploited, including **TensorFlow**, **PyTorch**, **Keras.io**, **MLlib**, as they were defined on Section 8.5 of the current study. For the model serving phase, respectively, **TensorFlow serving**, **Torch Serve**, **KFServe**, are promising candidate solutions. On this direction, Lot1 suggests the adoption of **TensorFlow** and **TensorFlow serving** frameworks, as the most appropriate solutions for the current scenario of **BUC.02**, since both tools are widely exploited in industry and in production environments, due to their **high-flexibility**, **high-performance**, and **speed** in training and serving **ML/AI models**. Moreover, for the code developing/writing part, Lot1 suggests the adoption of interactive notebooks solution, in which several users can be connected and write their incremental developments. Characteristic examples of several well-exploited interactive notebooks, as investigated on Section 8.5 are: **JupyterHub**, **JupyterLab**, and **KubeFlow's** notebook. Finally, **Lot1** suggests the adoption of **JupyterLab** solution, since it enables users to work with documents & activities such as Jupyter notebooks, text editors, terminals, and custom components in a flexible manner, and incorporates interaction within other interactive notebooks, such as **JupyterHub**.

To conclude, for the data visualization part, **Lot1** suggests as candidates the tools posed in Section 8.6.

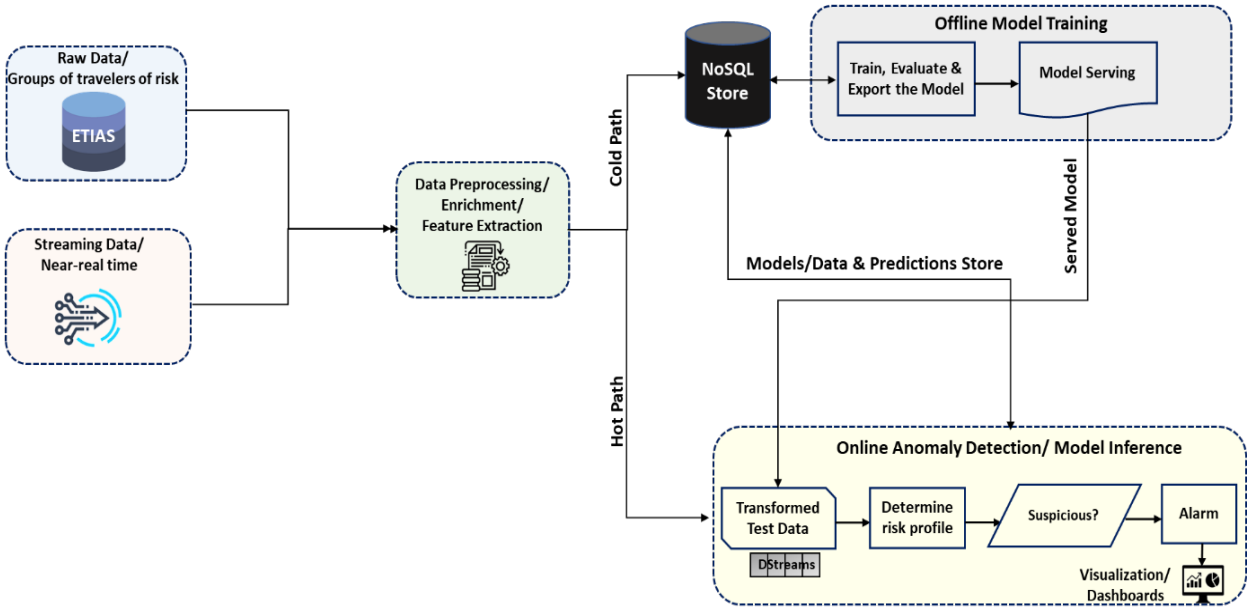


Figure 30. Proposed High-Level Flow Diagram for BUC.02

5.4.3 BUC 03 - Analyse Risk Profiles Correlations

BUC-03

Name	ANALYSE RISK PROFILES CORRELATIONS
Description	The purpose of this business use case is to enhance the **ETIAS** analytical capabilities beyond the deterministic approach, by using AI technology to **support human decision making**. The AI technology will identify and highlight correlations amongst the risk profiles from its **historical data** during the **ex-ante assessment** allowing a more precise definition.
Data Structure/ Dataset	Structured/ Semi-structured and Anonymized data, including attributes extracted from: **ETIAS**, including: { Status Information", "Nationality", "Sex", "Year of Birth", "Country of Residence , "Education", "Current Occupation", "Type of Travel Document", "Three-letter code of the issuing country", "Type of Travel Authorization", "Limited Territorial Validity", "Validity Period for Travel Authorization", "Refusal Reason"}. **EES**, including: {"Status Information", "Nationality", "Sex", "Year of Birth", "Date of crossing", "Border crossing point", "Type of Travel Document", "Three-letter code of the issuing country", "Number of Persons identified as over-stayers", "Nationalities of persons identified as over-stayers", "Revoked Stay", "Validity Extension", "Fingerprints", "Refused Entry", "Refused Entry Nationality", "Refused Entry Type of border", "Refused Entry border crossing point", "Refusal Reason"}. **WHO & ECDC epidemiological reports.**[54]
AI Methodology	**BUC.03** exploits novel formulations to define: (i) risks related to security or illegal immigration, and/or (ii) high epidemic risks, between ETIAS, EES, and reports provided by ECDC and WHO, when applicable. Correlations on the **abnormal rates of overstaying/ refusals of entry**, for **specific groups of travellers** among **ETIAS** and **EES** will be extracted and will further demonstrate the interconnectivity of the different systems. Additionally, based on the provided information regarding epidemiological reports provided by ECDC and WHO, and ETIAS/ EES systems, correlations indicating high epidemic risks can also be extracted. To achieve the above-mentioned objectives, the proposed **AI Solution** relies on **Pattern Recognition approaches**, and is further responsible for highlighting the amount of correlation between the ETIAS/ EES risk profiles. Additionally, the proposed **AI Solution** will demonstrate how the results from the grouping process into groups of travellers from the different systems are connected to each other. The key benefit of using AI lies into the fact that it can **uncover correlations** between the already defined **risk-profiles**. Both **Semi-supervised** and **unsupervised** Pattern Recognitions approaches will be considered for **BUC.03**.

[54] In the current analysis, we consider risk-profiles correlations among EES and ETIAS system, while in the near-future, reports from WHO&ECDC can be also considered and incorporated (the main design framework of BUC 03 will not be changed)

BUC-03

Indicative Implementation Steps	**Indicative key steps of the implementation process are described below:** • Data pre-processing for the risk profiles that are extracted from the various data sources, including processing of structured and semi-structured information & further data selection. • Data cleansing and re-organization to obtain the required attributes (i.e., columns/ analysis variables). • Setup the parameters for the different correlation clustering algorithms. • Execute the correlation clustering algorithms. • Generate correlation clusters. • Perform cluster quality for the proposed algorithms based on certain metrics. **Process Output:** Correlation clusters indicating the **amount of similarity or divergence of the investigated risk profiles**.
Key Success Factors/ Pre-requisites	**Key Success Factors/ Pre-requisites for BUC.03 may be considered the following:** • Enough ETIAS and EES data to perform the analysis. (The reader may refer to the corresponding stakeholder requirements from the Section Stakeholder Requirements (RP.03). • Reports from WHO and ECDC indicating epidemic risks • Definition of the WHO/ ECDC report's structure • Definition of which attributes/ keywords are of interest to be automatically extracted from WHO/ ECDC reports. • Reflect the corresponding requirement for preventing direct or indirect discrimination in the application of the datasets, i.e. REQ-50.01 (BR): The CRRS shall not receive or process any personal data that could allow discrimination against persons on any grounds or violating human dignity, integrity and fundamental rights.
Assumptions	**Key assumptions for BUC.03 may be considered the following:** All data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved.

5.4.3.1 Description of the Best Fit Solution

The purpose of this Business Use Case (**BUC.03**) is to uncover **similarities** and/or **regularities** between different **risk profiles** generated with data coming originally from **ETIAS** or **EES**. According to the stakeholder requirements, the correlations are related to **abnormal rates of overstaying/refusals of entry** for specific groups of travellers belonging both to **ETIAS** and **EES**. Furthermore, correlations indicating **high epidemic risks** utilizing information coming from epidemiological reports generated by ECDC and WHO, and ETIAS/EES systems can be considered and revealed as well (i.e., however this is a part of a future study in which we will obtain concrete knowledge of the precise attributes that need to be derived from such reports). Thus, the dependencies between different systems will be revealed and lead to useful conclusions regarding the created risk profiles and how this procedure can be further enhanced.

The key contribution of **BUC.03** is to comprehend the **patterns between risk profiles**, and further **understand any underlying behaviours behind them** or **improve the classification process** to lead to more **tangible**, **robust** and **reliable** results. For this purpose, **Pattern recognition** algorithms can be used in several different cases to identify any hidden trends and regularities[55]. The recognition of trends can be done either physically by human interaction or using state-of-the-art **Machine Learning** techniques[56]. However, the main scope of the current study is to highlight how **AI shall contribute with human intelligence to produce more efficient and non-biased risk profiles**.

[55] Zhang, X Y , Liu, C L , & Suen, C Y (2020) Towards robust pattern recognition A review Proceedings of the IEEE, 108(6), 894-922
[56] Abiodun, Oludare Isaac, et al "Comprehensive review of artificial neural network applications to pattern recognition " IEEE Access 7 (2019) 158820-158846

Pattern recognition approaches has multiple applications in various fields, for example speech recognition, speaker identification, image processing, computer vision, seismic analysis, and text analysis, among others. To serve the needs and requirements of **BUC.03**, the pattern recognition approach that will be included in this study relies on **text analysis**. Text analysis is used to analyse and compare segments of text. Thus, patterns shall be used to "**understand**" textual information and perform analysis on it. Moreover, text recognition is used to classify documents as well as extract sensitive information from them in an automated manner, which will be enriched though with human interventions (i.e., eu-LISA's experts). On this direction, similarity measures for text clustering will also be considered[57].

Figure 36 illustrates a high-level way flow diagram of **BUC.03**. First, data is inserted mainly by **ETIAS** system and on second note on by other systems, i.e., EES in this scenario. There are two paths in which data is inserted; either **hot path** or **cold path**. Hot path is for data that are captured in *near-real time*, whereas cold path refers to data that are going to be stored in the data store and considered *historical data*. Consequently, the raw data will be transferred into a data processing step, where data will be transformed into a suitable format for analysis purposes. The data of the cold path, as mentioned earlier, is stored into a proper data store. The predominant choices are: **Apache Cassandra database**, or **MongoDB**, or **Redis**. In this **BUC,** the need of a **NoSQL** data store is undeniable since the type of data that will processed need special attention and flexibility capabilities.

Historical data is going to function as the training dataset and the respective algorithms are going to be created. This procedure constitutes the **offline model training phase**. For this phase (**ML/AI training phase)** of the current **BUC** the tools that were selected were **TensorFlow**, **PyTorch**, **Keras.io** and **MLlib.** However, **Tensorflow** is considered a better candidate to meet the needs of **BUC.03**. Additionally, for the coding developing part, and the framework that is proposed by **Lot1** is **Jupyter Lab** and particularly, **Jupyter notebook**, for handling the data modelling phase. In addition, for the model serving phase the possible candidates are **TensorFlow serving**, **Torch Serve**, and **KFServe.** Since **Tensorflow** is proposed, **TensorFlow serving** will match the solution better and offers an alignment. Furthermore, streaming data are used in the online testing process that validates the model's robustness, i.e., the **extracted patterns/ model inference phase**. The metadata mechanism will operate similarly on both processes (**hot path**, **cold path**). The results of the model shall be stored in the main **NoSQL** database. Last, the model is served and utilised for virtualisation purposes if applicable. Data Visualization tools are discussed extensively in section **7.6**.

[57] https //www cscjournals org/manuscript/Journals/IJDE/Volume2/Issue4/IJDE-63 pdf

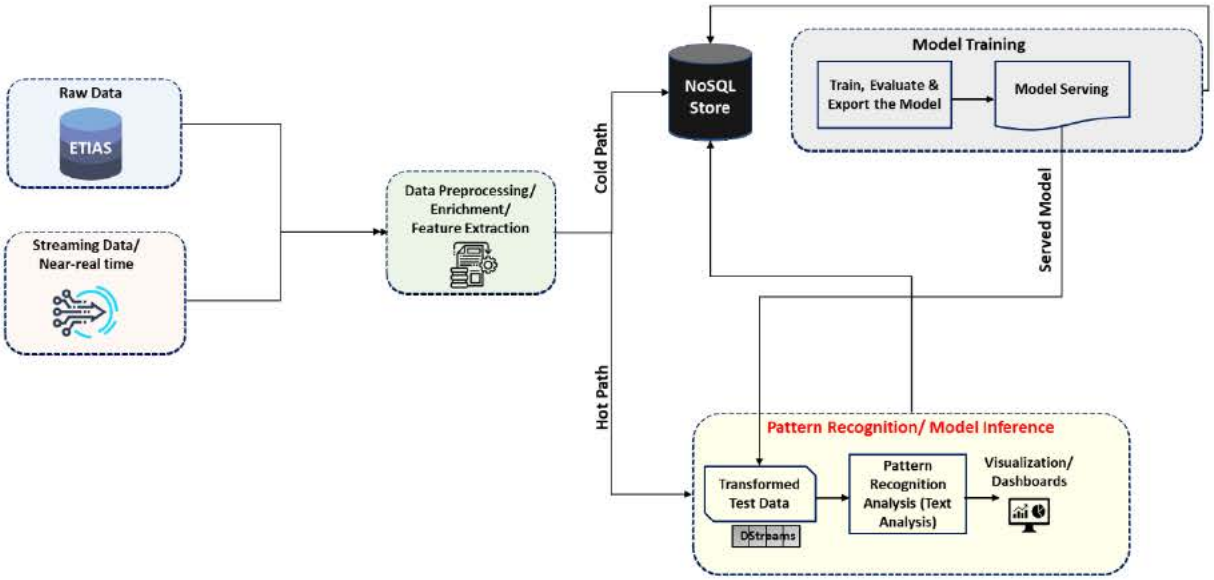


Figure 31. Proposed High-Level Flow Diagram for BUC.03

5.4.4 BUC 04 - Automatic Processing of Textual Information from Various Sources

BUC-04

Name	AUTOMATIC PROCESSING OF TEXTUAL INFORMATION FROM VARIOUS SOURCES.
Description	The purpose of this business use case is to enhance the CRRS analytical capabilities beyond the deterministic approach, by using **AI technology** to support human decision making. Once the respective source of information (i.e. please refer to the field below) is manually uploaded in CRRS by a Member State (MS) or by any other duly authorised staff, **AI technology** will identify text patterns or a set of common textual characteristics from the analysis of data allowing the identification of a group of travellers of risk (Clusters) that meets the criteria for the validation of a proposed risk profile.
Data Structure/ Dataset	Text information (i.e., reports) from the various CBSs that reach CRRS (i.e., ETIAS, VIS, EES), particularly datasets as part of system record extraction, containing textual information.
AI Methodology	In this **Business Use Case (BUC)**, we propose a robust **AI Solution** able to process and further understand, in near real-time conditions, **text information**. Specifically, **BUC.04** covers the scenario where text reports reach CRRS, and further contain information able to allow the **identification of a group of travellers of risk** that meets **certain criteria** for the **validation of specific risk profiles**. Consequently, in **technical terms**, the **main objective** of BUC.04, is to identify **text patters/ keywords of interest**, and **segments** (i.e., **parts**) of **text**, which enable the identification of groups of travellers of risk for certain risk profiles. Key difference among BUC.01, BUC.03 and BUC.04, regards the data type, since in the current BUC, **Lot 1**considers only text information. **Human intervention** is crucial towards the extraction and verification of the proper keywords of interest, or the part of the documents that are of special interest for eu-LISA's stakeholders, from various text reports that reach CRRS. By adding an extra level of intelligence, **Deep Learning** approaches, will be able to provide better insights on the specific **BUC**. On this direction, **Deep Learning Classification models**, such as: **Convolutional Neural Networks (CNNs)**, **Recurrent Neural Networks (RNNs)**, **Long Short-Term Memory Networks (LSTMs)**, for **text sequence processing** can be used to further categorise documents, among other methodologies that will be also studied in future studies. Finally, the trained models will be also able to learn from patterns in the already existing (i.e., historical) text reports.

BUC-04

Indicative Implementation Steps	**Indicative key steps of the implementation process are described below:** Collect the reports from the different CBSs (i.e., VIS, EES, ETIAS, etc.). • For each text report, extract meta-data information (i.e., text pre-processing step), including: tokenization, removing stop words, stemming and vector space models generation. Tokenization is used to remove the white space and special characters, while stop words removal is implemented to not include information that carries little meaning. Stemming is used to remove the suffixes and prefixes from keywords. It is the process of identifying the stem word or the root word of a set of words which share the same stem etymologically. This will reduce the vocabulary size, leading to better and more robust representations of the input dataset. Finally, vector space models represent each term in the learning group as a vector composed of a certain number of dimensions multiplied by a certain number of terms. • Perform feature extraction to convert text from any kind of format or setup, into a well-defined keyword schedule which it may be easy to process for the subsequent learning process. • Perform dimensionality reduction via standard approaches, to decrease the computational resources that will be considered during the model training phase. • Split the dataset into training & validation sets using state-of-the-art practices. • Perform model training with different AI/ ML-based text classification algorithms to select the best possible algorithm, via a thorough evaluation approach. • For each AI/ML-based trained algorithm, extract the corresponding evaluation metrics indicating the robustness of the proposed AI technique. • The algorithm which ranks the highest evaluation metrics (i.e., **accuracy**), for both training and validation sets, will be selected for the text processing BUC. **Process Output:** The output of this analysis is expected to be the **required keywords of interest** for the **ETIAS** database.
Key Success Factors/ Pre-requisites	**Key success factors/pre-requisites for BUC.04 may be considered the following:** • Definition of the "target" keywords that need to be extracted via the proposed automatic/ semi-automatic text classification process (e.g., Status Information", "VISA Authority", "Location of Visa Authority", "Nationality", "Member State", "Date of VISA Application", "Place of VISA Application", "Type of VISA", "Type of Travel Document", "Number of VISA Applications", "Status of VISA Application", "Date of Acceptance", "Date of Refusal", "Purpose of Journey", "Fingerprints", etc.) – According to stakeholder requirement: "REQ-04.01.03 (SHR): The CRRS shall provide mechanisms to identify and extract keyword(s) from semi-structured or unstructured data." • All reports will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved. – According to stakeholder requirement: " REQ-04.01.02 (SHR): The CRRS shall ensure that all reports will be anonymized and no personal/sensitive information for the individuals is able to be retrieved."
Assumptions	**Key assumptions for BUC.04 may be considered the following:** • All data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved.

5.4.4.1 Description of the Best Fit Solution

The main purpose of **BUC.04** is to derive **text patterns** or a **set of common characteristics** from text reports in order to extract significant attributes that will be subsequently stored into CRRS. These attributes will be able to enhance all activities which lead to the identification of a group of travellers of risk (i.e. clusters) that fulfil the criteria of a certain risk profile, and by levereging on **AI capabilities**. As it is formulated within the stakeholder requirements, the

key concept is to utilize the information from CRRS that is selected from the various CBS sources (i.e. VIS, EES, ETIAS etc.) and produce valuable insights, such as the required keywords of interest for the ETIAS database.

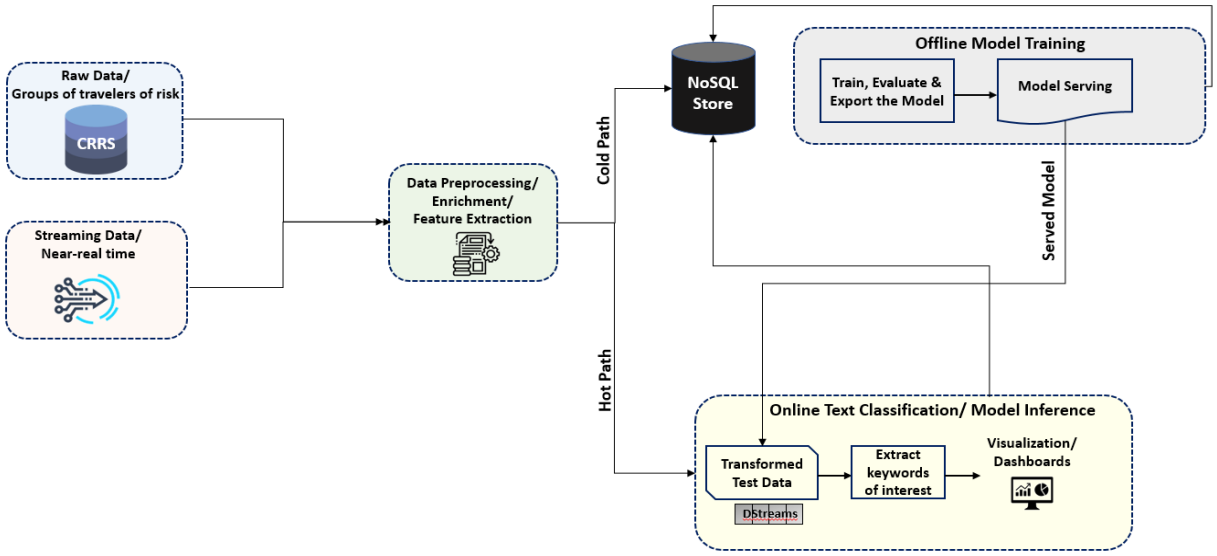


Figure 32. Proposed High-Level Flow Diagram for BUC.04

Key objective of the current BUC is to extract text patterns from any report landed into CRRS to identify groups of travellers of risk with certain criteria. These criteria are expected to be decided according to the Stakeholder's needs. Additionally, text patterns from individual Travel Authorization (TA) reports can also be considered on the specific BUC. **Figure 37** depicts the high-level flow diagram of the current BUC (**BUC.04**). As it was thoroughly explained in the previous use case (BUC.03), there are two types of data paths that will be followed when data is inserted in the system in order to perform the described analysis: **hot path** (i.e. **near-real time data**) and **cold path** (i.e. **batch data**). The tools for ingestion that **Lot1** decided to include in its study are **Apache Spark Streaming** for (near) real-time processing, while for batch data ingestion **Apache Flink** and **Apache Beam tools are considered**. All candidate tools for data ingestion are presented in detail on section 7.1. Once the data is ingested, the processing phase begins. The steps of processing include tasks such as **data pre-processing**, **enrichment**, and **feature extraction**, focused on **text data sources**. Specifically, within the text pre-processing steps, is considered the tokenization process, i.e., via removing stop words/ special characters, etc.; and stemming processing, i.e., by reducing inflection in words to their root forms such as mapping a group of words to the same stem. Additionally other text pre-processing techniques, such as lemmatization can be considered[58]. The pre-processing techniques, will significantly facilitate the reduction of the input data vocabulary, leading to more robust representations of the data, and consequently, into better ML/AI models.

Afterwards, the data of the batch process are stored for keeping track of the historical information into the **NoSQL** database. As it has already been mentioned in previous BUC.03, the NoSQL data stores that we concluded on, are **Apache Cassandra, MongoDB**, or **Redis.** Batch data is used for training purposes (i.e., **offline model training phase**), while the streaming data is be used for the **online text-classification,** forming the **model inference stage**.

All ingested data is proceeded into the corresponding data processing phase, which is considered a highly significant step to implement a robust data modelling process and produce reliable results. For the **ML/AI training phase** of the current BUC, the selected tools are: **TensorFlow**, **PyTorch**, **Keras.io** and **MLlib.** The total list of the

[58] https //nlp stanford edu/IR-book/html/htmledition/stemming-and-lemmatization-1 html

ML/AI tools that are considered can be found under Section 7.5. Furthermore, **TensorFlow serving**, **Torch Serve**, and **KFServe** are candidates for the **model serving phase**, as well. However, the selected frameworks are **TensorFlow** and **TensorFlow serving**, which are the most predominant candidates as they provide flexibility, high-performance and reliability. As part of the solution, **Jupyter Lab** is suggested and more specifically it is notebook component, **Jupyter** notebook. After the model has been through both phases, the **exact keywords of interest are extracted** and the results are either stored in the NoSQL data stores or **consumed** into **visualization tools**. A discussion about the visualization tools can be found in section 7.6 . In conclusion, the AI **solution**, includes all the necessary **technologies** and **tools** which are considered more suitable for its implementation, while it simultaneously involves **well-designed steps** that will enhance the procedure of the ***automatic extraction of keywords of interest from text reports that reach CRRS***.

5.4.5 BUC 05 - Ex-Post Assessment Process

BUC-05

Name	EX-POST ASSESSMENT PROCESS
Description	The purpose of this business use case is to enhance the ETIAS analytical capabilities beyond the deterministic approach, by using AI technology to support human decision making. The **AI Solution** will enhance the **ex-post assessment** process by **analysing** and **detecting deviations** and proposing the review of the risk indicators. This process can be triggered manually by a user or set within a **pre-scheduled timeframe** (e.g., **6-months window** or **smaller time-intervals**).
Data Structure/ Dataset	Structured and Anonymized data, including attributes extracted from **all CBSs** including: { Status Information", "VISA Authority", "Location of Visa Authority", "Nationality", "Member State", "Date of VISA Application", "Place of VISA Application", "Type of VISA", "Type of Travel Document", "Number of VISA Applications" , "Status of VISA Application", "Date of Acceptance", "Date of Refusal", "Purpose of Journey", "Fingerprints", "Sex", "Year of Birth", "Date of crossing", "Border crossing point", "Three-letter code of the issuing country", "Number of Persons identified as over-stayers", "Nationalities of persons identified as over-stayers", "Revoked Stay", "Validity Extension", "Refused Entry", "Refused Entry Nationality", "Refused Entry Type of border", "Refused Entry border crossing point", "Refusal Reason", "Country of Residence", "Limited Territorial Validity", "Validity Period for travel authorization"}.
AI Methodology	In this Use Case we design an **AI Solution** that **models the evolution of risk profiles with respect to a predefined time-period of 6-months**, but also **in smaller timeframes**. This problem can be assessed as a problem of "**Fraud/ Anomaly Detection**", in the sense that irregular patterns of risk profiles should be identified, during a pre-defined time-period. To tackle this BUC, we consider both the scenario where **labelled** and **partially labelled information**, exists (i.e., relying on historical data), and **Deep Learning Classification** methods will be designed to **categorize each risk profiles according to their validity**. Model training will be implemented on historical data of risk profiles, while the inference phase will extract a **score**, indicating whether the **risk profile** is "**valid"** or "**non-valid**". The outcome of this process will be decided based on pre-defined thresholds that will be agreed with eu-LISA stakeholders, and further adjustments can be considered, i.e., more output validity categories. Both **supervised** and **semi-supervised Deep Learning** approaches are considered, including**: Recurrent Neural Networks,** and the special category of the **Long-Short Term Memory Networks (LSTMs)**, both incorporating the **time-dimension** within the model training; Convolutional Neural Networks (CNNs) for extracting deep patterns on the input data; or combinations of the architectures.

BUC-05

Indicative Implementation Steps	**Indicative key steps of the implementation process are described below:** • Extract meta-data information from the various CBSs database (i.e., dataset enrichment/ feature extraction, etc.). • Agree with eu-LISA's stakeholders on the particular risk profiles that will be considered and labelled "**valid**". • Agree with eu-LISA's stakeholders on the particular risk profiles that will be considered and labelled "**non-valid**". • To generate the ML-based models, we split the dataset into training & validation sets using state-of-the-art practices. • Perform model training on the labelled "**valid**" and "**non-valid**" risk profiles using different ML-based algorithms and build the representative ML models. • Extract the corresponding evaluation metrics for both training & validation sets. • The algorithm which ranks the **highest evaluation metrics** and **accuracy**, for both training and validation sets, will be selected for the specific BUC. • For the testing phase, when a risk profile is marked for ex-post assessment, a label will be assigned based on the already trained model, indicating whether the risk profile is "**valid**" or "**non-valid**". **Process Output:** The output of this analysis is expected to be the "**valid**" or "**non-valid**" indication on the **investigated risk profile**.
Key Success Factors/ Pre-requisites	Key Success Factors/Pre-requisites for BUC.05 may be considered the following: • Precise definition of the analysis time-interval (e.g., 6-months or less) - According to the stakeholder requirement "REQ-05.02.03 (SHR): The CRRS shall allow a precise definition of the analysis time-interval (e.g., 6-months or less) for the ex-post assessment analysis." • All data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved. – According to stakeholder requirement "REQ-05.02.01 (SHR): The CRRS shall ensure that all data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved."
Assumptions	Key assumptions for BUC.05 may be considered the following: All data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved.

5.4.5.1 Description of the Best Fit Solution

The main goal of **BUC.05** is to define a well-structured **AI Solution** able to perform a comparative analysis of the already validated risk profiles performance (i.e., **ex-post assessment**). As it was explicitly described on the stakeholder's requirements on Section 9.10, the proposed **AI Solution** will be responsible for **validating risk profiles** based on their **ex-post assessment results**, according to the **identified patterns**, **trends** and **behaviours** of the **risk profiles**; while the results/outputs of this analysis will be exported to eu-LISA's authorised users. In more technical terms, the proposed AI methodology is responsible for monitoring the evolution of risk profiles within a predefined time-window of 6-months (i.e., the default), or even is smaller time-intervals, and extract indicators whether these risk profiles are "**valid**" or "**non-valid**". **Figure 38.** provides the proposed, high-level flow diagram of BUC.05, while the following paragraphs explain thoroughly Lot1 solution, along with the suggested AI methodology.

In the first step of the analysis, we exploit both: (i) **historical/ batch data** stored on the databases of the various CBSs, indicating abnormal rates of refusals of travel authorisations, and further corresponding to the groups of travellers or risk, and **near-real time/ streaming data** of risk profiles that require validation regarding their status. For the batch data case, we consider historical data extracted during the pre-defined time window of **6-months** (additionally smaller time windows can be examined- however the 6-months' frame provides a more holistic view for the analysis). The specific historical data will formulate our training dataset, and representative ML/AI models can be created using the corresponding data observations, formulating the **offline model training phase.** In

addition, streaming data will be used for the **online risk validity process,** i.e., the **model inference phase**. Both data sources, **historical** and **near real-time**, will be pre-processed via the same mechanisms to extract valuable metadata information and representative features, necessary for the **AI model training** and **inference** phases.

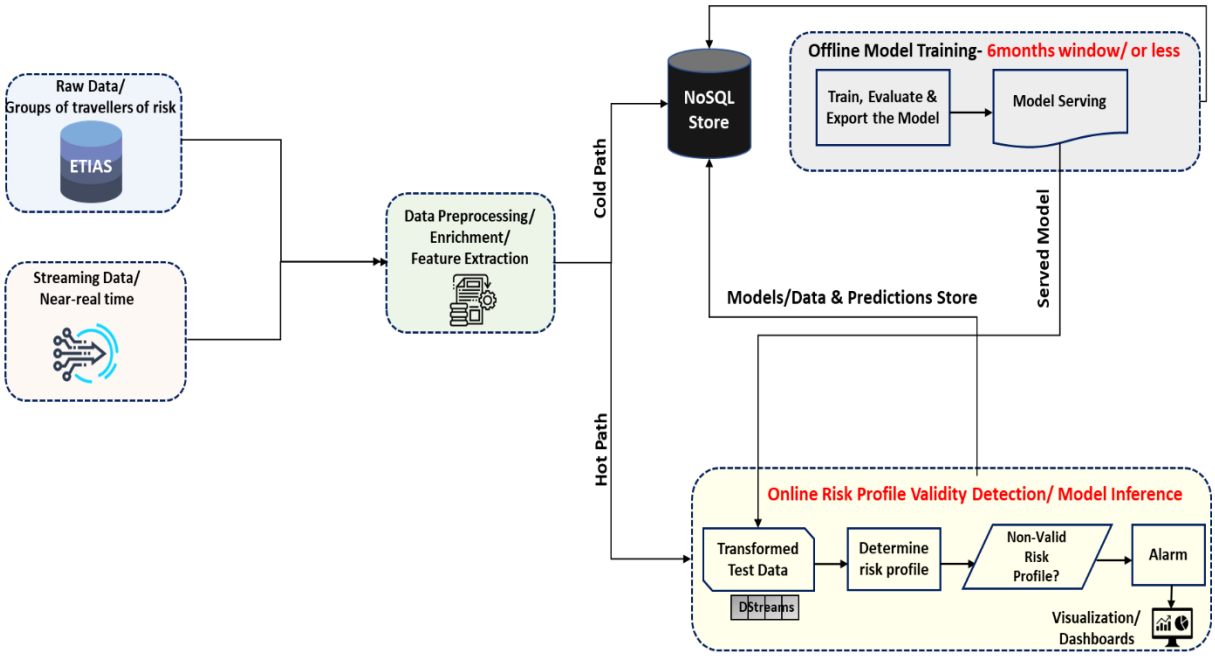


Figure 33. Proposed High-Level Flow Diagram for BUC.05

Processed batch/historical data from data pre-processing phase, are stored on a **NoSQL** database, to interact and be consumed by the model training component of the offline model training process. Specifically, the **model training** component holds all necessary **AI tools** and **algorithms** which are responsible for the ML/AI model training, the metadata extraction of this phase, along with the evaluation metrics from the training process, i.e. indicating the model's reconstruction errors, and the training dataset's accuracy, indicating the quality and robustness of the trained ML/AI model. Subsequent step to this process is the model serving to the **online risk validity detection** module, i.e., the **model inference phase**. In this step, the transformed test data, i.e., the examined risk profiles are combined with the served ML/ AI model, and extract probabilities for each different category, i.e., "**valid**", or "**non-valid**", the highest probability results into the main conclusion of the analysis, indicating the main inference. In case where a risk profile is characterised as "non-valid", according to the historical training analysis that was implemented, the result will be displayed on an interactive dashboard, in which eu-LISA's stakeholders and ETIAS authorised users could have access, and proper countermeasures from their side can be taken. Additionally, all results, of either "**valid**", or "**non-valid**" risk profile will be also stored into the **NoSQL** database.

Adhering to the paradigm of BUC.02, the investigated AI tools that may be adopted during the **ML/AI training phase** of **BUC.05** are **TensorFlow**, **PyTorch**, **Keras.io**, **MLlib**, as they were explicitly defined on Section 7.5. Respectively, **TensorFlow serving**, **Torch Serve**, and **KFServe** stand as excellent possible solutions to be utilised for the **model serving phase**, since they are widely exploited in several heavy-computation tasks along multiple industries, with high-efficiency. Specifically, Lot1 suggests the adoption of **TensorFlow** and **TensorFlow serving** frameworks, as the most appropriate solutions for **BUC.05**. For the code developing part, we find very convenient the opportunity of using interactive notebooks solution, while **JupyterLab** stands as a very good candidate that works with several documents & activities, including Jupiter notebooks, text editors, among others, enabling connectivity with other interactive notebooks. Last, for the data visualization part, Lot1 suggests as candidates all tools posed in Section 7.6.

5.4.6 BUC 06 - Virtual Assistance to Provide Guidance on Certain Activities

BUC-06

Name	**VIRTUAL ASSISTANCE TO PROVIDE GUIDANCE ON CERTAIN ACTIVITIES**
Description	The purpose of this business use case is to provide virtual assistance to the ETIAS Central Unit that responds to the contextual support/service/help. User interaction with virtual assistants will be implemented to **improve several critical business** processes, including **information on how to use the systems**, i.e., from the practical handbook (Art.93 from (EU) Reg 2018/1240), suggestions on **how to contact customer support teams** and to further **report incidents.**
Data Structure/ Dataset	Structured and Anonymized data, including attributes extracted from ETIAS including: { Status Information", "Nationality", "Sex", "Year of Birth", "Country of Residence , "Education", "Current Occupation", "Type of Travel Document", "Three-letter code of the issuing country", "Type of Travel Authorization", "Limited Territorial Validity", "Validity Period for Travel Authorization", "Refusal Reason"} These data are related with the purpose of the specific Use Case, if we consider for instance the scenario where a Chatbot requests for the Type of Travel Document from a travel applicant in order to facilitate in certain subsequent activities. Virtual Assistants can be used either by travel applicants or by the relevant authorized stakeholders in order to be provided with answers to certain questions about ETIAS. Additionally, similar methodology and design may also be extended and applied to other sub-systems as well.
AI Methodology	Chatbot solutions, Natural Language Processing/ Reinforcement learning/ Predictive Modelling / Supervised/ Semi- & Unsupervised Machine Learning Model Training according to the amount of labelled data have the potential to generate coherent answers to certain questions. The models could also learn from patterns in previous search behaviours.
Indicative Implementation Steps	**Indicative key steps of the implementation process are described below:** • Ensure that all required libraries, dependencies, and packages are installed; • Prepare ETIAS data on a proper format for the specific BUC; • Send a message to the ETIAS chatbot; • The proposed NLU process analyses the sentence and returns entities; • Return the possible intent of the message chatbot received by regular expression and keywords; • Respond to the message according to the intents and current states based on the state machine. • Validate the accuracy of the proposed approach, by calculating related metrics, such as: **intent accuracy** (i.e., correct recognized intents divided by the total number of intents), **entity accuracy** (i.e., the number of correct extracted entities divided by the total number of entities), **integrity of entity** (i.e., the number of recognized entities divided by the total number of entities), and the **integrity of sentence** (i.e., the total number of lines divided by the number of a sentence which entities are extracted completely). **Process Output:** The output of this analysis is expected to be the **answers** and **indicative actions**, proposed by the chatbot/ virtual assistant to be performed by the ETIAS authorised users.

BUC-06

Key Success Factors/ Pre-requisites	Key Success Factors/Pre-requisites for BUC.06 may be considered the following: • Precise definition on the data format, i.e., text. (According to stakeholder requirement: "REQ-06.01.02 (SHR): The CRRS shall include a precise definition on the text data format used as input for the chatbot.") • Precise definition on the coding language & required packages/ libraries. (According to stakeholder requirement: " REQ-06.01.07 (SHR): The CRRS shall include a precise definition on the coding language and required packages/libraries.") • Precise definition of the questions that need to be learned via the AI technologies. (According to stakeholder requirement: "REQ-06.01.06 (SHR): The CRRS shall include a precise definition of the questions that need to be learned via the AI technologies.") • Precise definition of the required answers/ actions that need to be extracted through the virtual assistant/ chatbot tool. (According to stakeholder requirement: "REQ-06.01.05 (SHR): The CRRS shall include a precise definition of the required answers/ actions that need to be extracted through the virtual assistant/ chatbot tool.") • Consider different variants of requests that are similar, and other variants that can be treated differently. (According to stakeholder requirement: "REQ-06.01.04 (SHR): The CRRS shall include different variants of requests that are similar, and other variants that can be treated differently.") Consider all ways a conversation can go. (According to stakeholder requirement: "REQ-06.01.03 (SHR): The CRRS shall provide a virtual assistant that contemplates all ways a conversation can go.")
Assumptions	**Key assumptions for BUC.06 may be considered the following:** • All data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved.

5.4.6.1 Description of the Best Fit Solution

Chatbots and virtual assistants commonly rely on heavy Deep Neural Network (DNN) computations, Natural Language Processing (NLP) technologies, Generative Models (i.e., such as Generative Adversarial Networks – GAN's), and Reinforcement Learning (RL). However, the main limitation of chatbots lies into their requirement of enormous amounts of training data to provide accurate results. On this direction, multiple libraries (i.e. libraries of the decided programming language that will be used for any AI/ML implementation) and pre-trained NLP-based models are implemented to cover the demands. Another limitation of such technologies regards the learning language. Certain languages, such as English, have a significant corpus of textual information that can be used to train neural networks, whereas other minor languages, might have limited corpus of data that can be used for training, which may negatively affect the quality of the algorithms. However, as virtual assistant's technology evolves, the above-mentioned limitations are continuously resolved. Taking into account the above-mentioned limitations and the great demand for adopting a chatbot solution for eu-LISA and their stakeholders, the main goal of **BUC.06** is to provide virtual assistance to the **ETIAS Central Unit** corresponding to the contextual support, service, and help. According to the requirements formulated for the specific BUC (i.e., "REQ-06.01.03 (SHR): The CRRS shall provide a virtual assistant that contemplates all ways a conversation can go." and "REQ-06.01.05 (SHR): The CRRS shall include a precise definition of the required answers/ actions that need to be extracted through the virtual assistant/ chatbot tool."), a chatbot-like solution will enable a natural-language search engine to find targeted answers to certain user's questions. For this purpose, Lot1 considers the adoption of **Rasa Open-Source** tool, providing all necessary building blocks for creating robust virtual assistants. **Rasa Open Source** is exploited to automate human-to-computer interactions anywhere from websites to social media platforms, while it is widely used by several organisations. **Rasa** consists of two main components: **Rasa NLU** and **Rasa Core**. Specifically, **RASA NLU** is the interpreter who processes ETIAS input information, identifies the intents, and extracts the required entities from it. On the other hand, **Rasa Core** receives the output from the **Rasa NLU** component (i.e., which is formulated as a dictionary that specifies the desired entities and other related metadata information). Based on these details, the **Rasa Core** component selects the appropriate reply and sends it back to the ETIAS users as the chatbot's reply.

The main goal of **BUC.06** is to provide virtual assistance to the **ETIAS Central Unit** corresponding to the contextual support, service, and help. According to the requirements formulated for the specific BUC, a chatbot-like solution will enable a natural-language search engine to find targeted answers to certain user's questions. For this purpose, Lot1 considers the adoption of **Rasa Open-Source** tool, providing all necessary building blocks for creating robust virtual assistants. **Rasa Open Source** is exploited to automate human-to-computer interactions anywhere from websites to social media platforms, while it is widely used by several organisations. **Rasa** consists of two main components: **Rasa NLU** and **Rasa Core**. Specifically, **RASA NLU** is the interpreter who processes ETIAS input information, identifies the intents, and extracts the required entities from it. On the other hand, **Rasa Core** receives the output from the **Rasa NLU** component (i.e., which is formulated as a dictionary that specifies the desired entities and other related metadata information). Based on these details, the **Rasa Core** component selects the appropriate reply and sends it back to the ETIAS users as the chatbot's reply. Main reasons that enhance our selection of the **AI-enabled tool of Rasa Open Source**, besides the fact that the specific solution is widely exploited[59],[60] are the following:

- its core component is Open Source and free;
- enables to incorporate many external state-of-the-art Natural Language Understanding (NLU)/ Natural Language Processing (NLP) models;
- it is highly customizable and flexible;
- it is core API relies on Python's Scikit-learn and Keras.io libraries;

[59] https //iopscience iop org/article/10 1088/1757-899X/1077/1/012060/pdf
[60] https //arxiv org/ftp/arxiv/papers/2204/2204 09717 pdf

- contains an easy-to-use command-line interface for common tasks such as models training and testing;
- contains well-defined and rich documentation, both in written/ technical documents and video tutorials.

In technical terms, we exploit data extracted from the **ETIAS** system, which are first directed into the component responsible for pre-processing, enrichment, and feature extraction and preparation for the Rasa Open-Source tool.0

The pre-processed data are also stored for future usage on a **NoSQL database**. Respective to the solutions of the other BUCs, Lot1 suggests the selection of either an Apache Cassandra, or MongoDB, or Redis database for storing semi-structured/ unstructured information as the extracted entities from the raw data processing part, along with the results from the chatbot's conversations. Model training component is responsible for training the chatbot to find the intent and entity of users' utterances. When the users enter their queries/ questions, the chatbot identify both the intent and the corresponding entities and retrieves the appropriate response from the knowledge base. Specifically, Rasa will call an endpoint that can be specified when a custom action is predicted. This endpoint should be a web server that reacts to this call, runs the code and optionally returns information to modify the dialogue state. All conversations will be tracked and stored in the NoSQL database, while simultaneously reports can be also generated and visualized in any of the suggested data visualization tools that were also exploited on the above-mentioned BUCs (i.e., **Pentaho** and **Superset**). The following figure illustrates the high-level flow diagram of BUC.06.

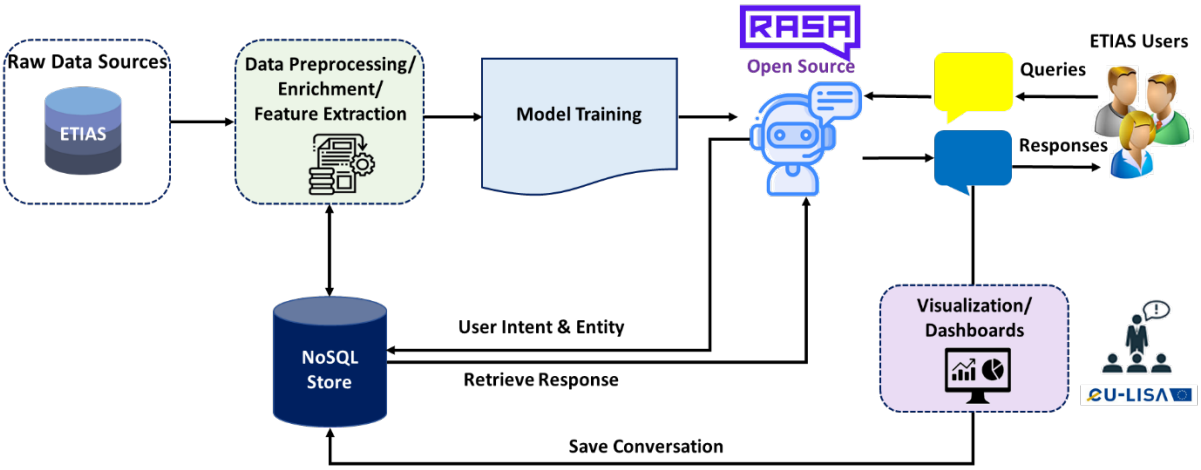


Figure 34. BUC.06 Proposed High-Level Flow Diagram for BUC.06

Regarding the process flow, the users enter data (i.e., queries) into the chatbot in natural language, which first needs to be tokenized, and then converted into numerical representations (i.e., feature extraction). Techniques that could be utilized are the **word2vec transformation**[61], and the **Continuous Bag of Words (Bow)**[62], among others, converting words/ sentences into vectors. Since the words are converted into numerical representations, the next step regards the intent classification, which is a sub-category of the general text-classification process. Intent/ text classification is the content understanding from the input Natural Language, provided by the users. Lot1 suggests being tackled as a classical problem of Machine Learning/ AI classification, as it was explicitly described on **BUC.04 – Automatic Processing of Textual Information from Various Sources**. Specifically, traditional SVM classification approaches can be used, however Lot1 proposed the exploitation of more sophisticated, **Deep Learning (DL)** formulations, including **bidirectional Long-Short Term Memory (LSTM) networks**[63], and **Convolutional Neural Networks (CNNs)**[64], among others. Even though various (text-related) pre-trained models are already built on RASA NLU component, in case where new models need to be trained and deployed, Lot1 suggests the exploitation of **TensorFlow**, **Porch**, and **Keras.io**, for the model training part, and the corresponding **TensorFlow Serving**, **Torch**

[61] https //www tensorflow org/tutorials/text/word2vec
[62] https //ieeexplore ieee org/abstract/document/7966208
[63] https //direct mit edu/neco/article-abstract/31/7/1235/8500/A-Review-of-Recurrent-Neural-Networks-LSTM-Cells
[64] https //www sciencedirect com/science/article/abs/pii/S0031320317304120

serve and **KFServe**, as possible candidates for the model deployment part.

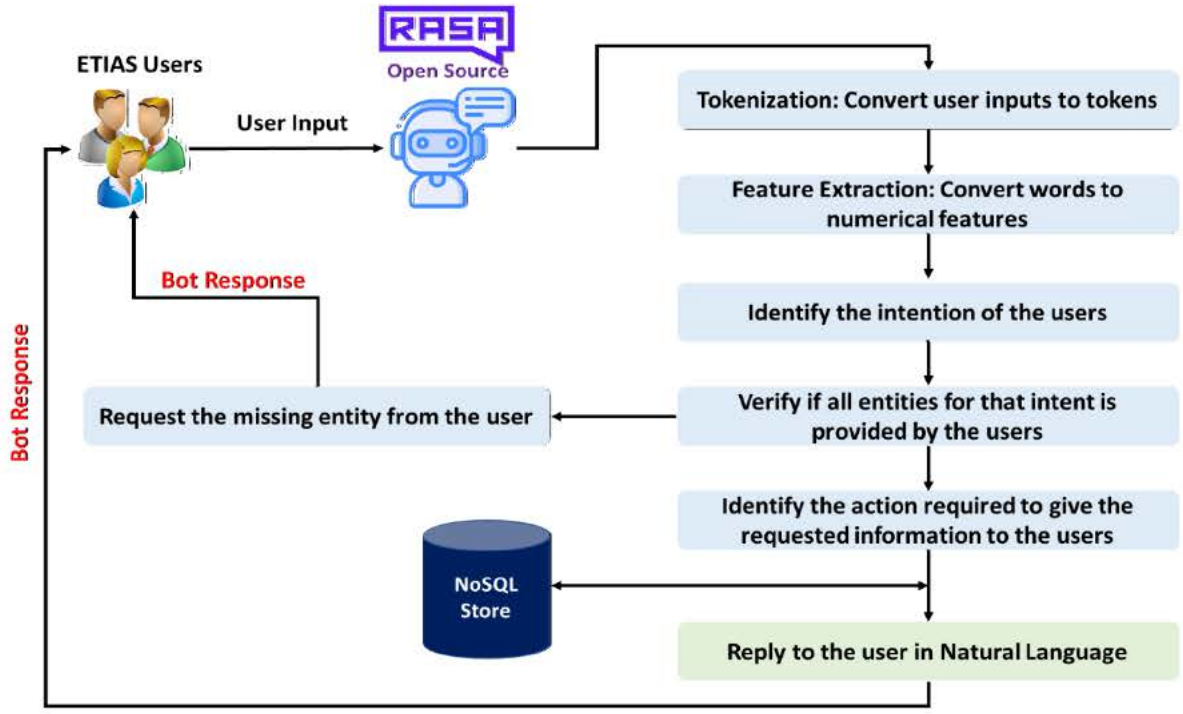


Figure 35. BUC.06 Proposed High-Level Process Flow

We should mention that the above-described steps of tokenization, feature extraction, intent classification, and entity extraction can be implemented via **Rasa Open-Source tool** using the **Rasa NLU pipeline**, while any step that requires dialogue management will be handed through **Rasa Core policies**. Finally, custom actions (i.e., data retrieval from the NoSQL database), and integrations with external APIs may be implemented via **Rasa SDK**[65].

5.4.7 BUC 07 - Verification process when a hit takes place

BUC-07

Name	**VERIFICATION PROCESS WHEN A HIT TAKES PLACE**
Description	The purpose of this business use case is to **support human decision making** of the ETIAS CU by applying **AI methodology** during the **verification process when a hit takes place** (*please refer to (EU) Regulation 2018/1240 Article 22(3) for more details about the verification process*).
Data Structure/ Dataset	Structured and Anonymized data, including attributes extracted from **ETIAS** including: { Status Information", "Nationality", "Sex", "Year of Birth", "Country of Residence , "Education", "Current Occupation", "Type of Travel Document", "Three-letter code of the issuing country", "Type of Travel Authorization", "Limited Territorial Validity", "Validity Period for Travel Authorization", "Refusal Reason"}

[65] https //package wiki/rasa-sdk

BUC-07

AI Methodology	In this Use Case we design an AI methodology that analyses and further compares ETIAS data to specific risk indicators to identify possible "hits". We are not directly comparing risk indicators with possible hits; we rather try to identify all relevant patterns within the provided data that will eventually identify "hits". Similar to BUC.02 and BUC.05, we tackle the specific BUC, as a characteristic scenario of **Machine Learning (ML)-based Anomaly/Fraud Detection**. Specifically, we consider both the scenario where labelled and unlabelled information, exists (i.e., relying on ETIAS historical data), and Deep Learning Classification methods will be designed to categorize ETIAS data/risk profiles according to their probability of producing a "**hit**" or "**no-hit**". Model training will be implemented on the historical data of the corresponding **risk profiles**, while the inference phase will extract a score, indicating whether the risk profile is producing a "**hit**" or "**no-hit**". This process will be decided based on pre-defined indicators related with the existence of "**hits**" in common agreement with eu-LISA stakeholders. Supervised, and unsupervised Deep Neural Networks (DNNs) approaches will be considered, including Recurrent Neural Networks (RNNs), Long-Short Term Memory Networks (LSTMs), Convolutional Neural Networks (CNNs), etc. or combinations of the architectures will be examined.
Indicative Implementation Steps	**Indicative key steps of the implementation process are described below:** • Extract meta-data information from the ETIAS database (i.e., dataset enrichment/ feature extraction, etc.). • Agree with eu-LISA's stakeholders on the characteristics of the ETIAS risk profiles that are most likely to produce hits, and further label them as "**hit-candidates**". • Agree with eu-LISA's stakeholders on the characteristics of the ETIAS risk profiles that are not likely to produce hits, and further label them as "**no-hit-candidates**". • To generate the ML-based models, we split the dataset into **training** & **validation** sets using state-of-the-art practices. • Perform model training on the labelled "**hit-candidates**" and "**no-hit-candidates**" ETIAS risk profiles using different ML-based algorithms and build the representative ML models. • Extract the corresponding **evaluation metrics** for both training & validation sets, i.e., **model accuracy**, etc. • The algorithm which ranks the **highest evaluation metrics** and **accuracy**, for both training and validation sets, will be selected for the specific BUC. • For the testing phase, when new ETIAS risk profiles are generated, a label will be assigned based on the previously trained model, indicating whether the specific risk profile is considered a "**hit-candidate**" or "**no-hit-candidate**". **Process Output:** The output of this analysis is expected to be the "**hit-candidates**" or "**no-hit-candidates**" indication on the **ETIAS incoming groups of travellers**.
Key Success Factors/ Pre-requisites	Key Success Factors/Pre-requisites for BUC.07 may be considered the following: • Precise definition of the historical analysis time-window. (According to the stakeholder requirement: "REQ-07.01.03 (SHR): The CRRS shall include a precise definition of the historical analysis time-window for hit classification.") • Agreement with eu-LISA's stakeholders on the anomaly detection levels (i.e., "**hit-candidate**", "**no-hit-candidates**") corresponding to the ETIAS risk profiles. (According to the stakeholder requirement: "REQ-07.01.02 (SHR): The CRRS will include anomaly detection levels corresponding to the ETIAS risk profiles and as agreed among eu-LISA's stakeholders.") • Enough ETIAS historical data to perform the analysis. (According to the stakeholder requirement "REQ-07.01.01 (SHR): The CRRS shall ensure enough historical data to support human decision-making during verification process on hits.")

BUC-07

Assumptions	**Key assumptions for BUC.07 may be considered the following:** All data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved.

5.4.7.1 Description of the Best Fit Solution

The goal of BUC.07 is to verify the scenarios whether "**hits**" takes place on ETIAS data. For this purpose, in the current assessment we compare ETIAS data to specific risk indicators, that will be commonly agreed with eu-LISA's (and more specifically) ETIAS stakeholders/experts. **Human intervention** is highly significant to the specific BUC, to provide the proper guidelines to the AI architects and data scientists, leading to the best possible solution for this challenging task. Similar to **BUC.02** and **BUC.05**, one of the initial design steps towards the definition of the architecture solution, and the subsequent implementation steps, is the generation of a high-quality training dataset, including historical observations, i.e. batch data, from ETIAS risk profiles that result either into "**hit-candidates**", when they are combined with certain risk indicators, commonly decided with eu-LISA's stakeholders, and "**no-hit-candidates**", when no risk/ or any other critical situation is identified.

For this purpose, AI architects/ data scientists and ETIAS stakeholders should be able to build a dataset, including observations where ETIAS risk profiles can be considered as "**hit candidates**", and the contradictory scenario, of ETIAS risk profiles where can be characterised as "**no-hit candidates**". BUC.07 is tackled as a characteristic scenario of **ML-based data classification/ anomaly detection**, in which we propose a robust mechanism that identifies the "**hit**"-status, using historical learnt observations of ETIAS risk profiles that result into "**hits**", and of ETIAS risk profiles that are **not reported as** "**hits**". The following diagram illustrates the proposed high-level flow diagram of BUC.07.

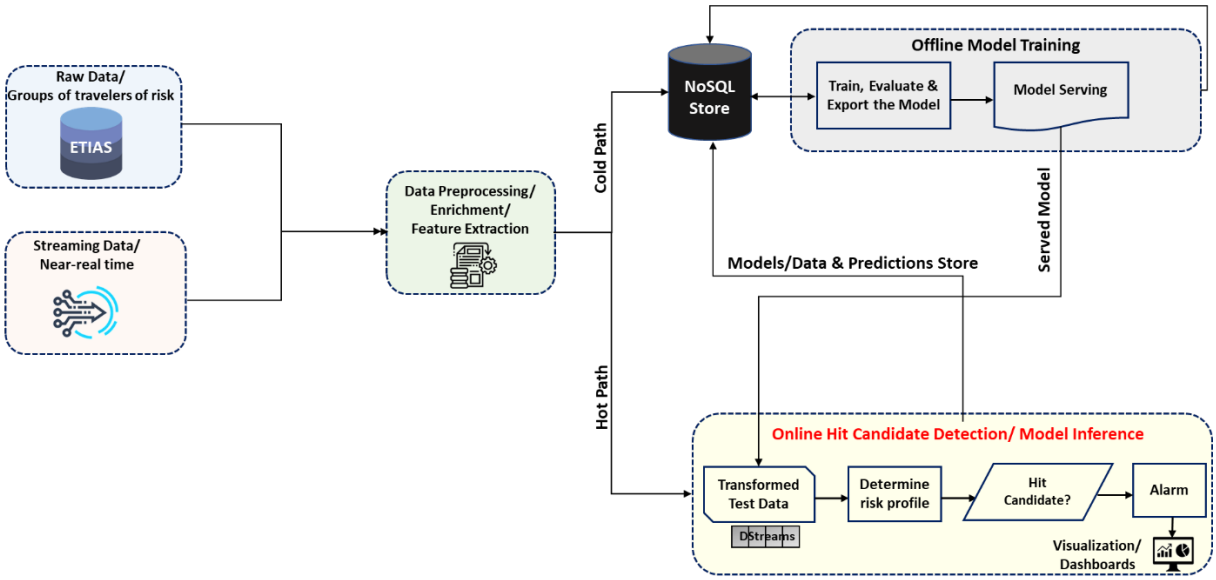


Figure 36. BUC.07 Proposed High-Level Flow Diagram

Both data types, i.e. historical ETIAS risk profiles and the upcoming streams of the new risk profiles are first passed through the **data pre-processing**, **enrichment** and **feature extraction** component, in which specific processing for the model training phase, takes place. On this direction, we should mention that data pre-processing is a very significant step towards the subsequent offline model training stage. Like the aforementioned BUCs, and specifically **BUC.02** and **BUC.05** that present several similarities towards the nature of the problem they are trying to solve, we set a strong prerequisite of the enormous increase of data, and the probability of having low query-time from eu-LISA's systems, and specifically ETIAS. For this purpose, as a Consortium, we suggest the exploitation of **Apache Spark Streaming** as the most appropriate tool for **(near) real-time processing**. Similar tools to Apache Spark that can be exploited, are deeply studied and provided within section 8.1, are **Apache Flink** and **Apache Beam**.

Historical data are stored in the **NoSQL** database component which interacts with the **model training**, **evaluation** and **export** component. The selection of the **NoSQL** database is highly efficient, since we can store several types of data, including ML models, metadata information extracted from the model training phase, and inference/ prediction results from the model inference phase. From the pre-processing and store step, we proceed then to the main learning framework, including the component which is responsible for designing (and subsequently) developing the **ML/AI** algorithms for **hit candidate detection**. This learning framework is separated into two distinct categories: (i) the **offline model training**, referring to all actions responsible for designing and developing robust **ML/AI** algorithms, along with the models' export and serving process; and (ii) the **online hit candidate detection**, referring to the inference phase, i.e., decision making stage. Regarding the offline-training stage, Lot1 suggests storing the learnt models on either an **Apache Cassandra** database, a **MongoDB**, or a **Redis** server, to be able to use them and distribute them for the final stage of the inference extraction. All options suggested by Lot1, after the thorough study and benchmarking that was performed and discussed in the respective Data Store section.

Concerning the online hit candidate detection phase, the server model from the offline phase is combined with the testing data, i.e., the upcoming ETIAS observations that reach the system, to determine the updated risk profile, and encode the probability of resulting into a "**hit-candidate**" or a "**no-hit-candidate**". The highest probability determines the final decision/ inference for the investigated risk profile. In the scenario where the risk profile is determined as "**hit-candidate**", Lot1 suggests being displayed on an interactive dashboard, able to be monitored by the responsible authorities (i.e., ETIAS, eu-LISA's authorized users). All results, of either "**hit candidates**", or "**no-hit-candidates**" will be also stored into the **NoSQL** database.

Like BUC.02 and BUC.05, the investigated AI tools that may adopted during the **ML/AI training phase** of the current BUC are **TensorFlow**, **PyTorch**, **Keras.io**, **MLlib**, as they were explicitly defined on Section 7.5. Likewise, **TensorFlow serving**, **Torch Serve**, and **KFServe** stand as excellent candidates to be adopted for the **model serving phase**. Due to their high-usability, performance, flexibility, and speed in training and serving ML models, Lot1 suggests the adoption of **TensorFlow** and **TensorFlow serving** frameworks, as the most appropriate solutions for BUC.07. Regarding the code developing process, as a Consortium we suggest the adoption of interactive notebooks solution, like all BUCs, and we state that **JupiterLab** is a very good candidate that works with several documents & activities, including Jupiter notebooks, text editors, among others, while is incorporates connectivity with other interactive notebooks. To summarize the current BUC, for the data visualization part, Lot1 suggests as candidates the tools posed in Section 7.6.

6 Chapter 6 – Conclusions and Recommendation

In this chapter we provide a reflection of the work that was implemented in this study and main recommendations on how eu-LISA (CRRS, ETIAS, and Recast VIS) and the proper Stakeholders, i.e., Frontex, and Commission can evolve their AI journey. Lot1 provided an **end-to-end analysis** and **assessment** on the **core concept** of this project, which is to "*identify meaningful Business Use Cases (BUCs), specific business and stakeholder requirements for the adoption of novel Artificial Intelligence (AI) formulations in CRRS, and specifically in the context of ETIAS and Revised VIS, with ultimate goal the support of the risk profiling, screening rules and risk indicators*". The business opportunities that were identified, the designed roadmap, along with the main enablers from the AI perspective, stand to help eu-LISA towards a more flexible journey towards critical decision making. Despite AI is designed to be the core concept of this project, **human intervention** is the **key indicator** towards the **success of this project**. AI will not be able to automatically ensure non-discrimination. However, AI's trustworthiness, such as the concepts of data protection and non-discrimination, will have to be agreed with eu-LISA and its stakeholders. In the context of AI, equality entails that the system's operations cannot generate unfairly biased outputs (e.g. the data used to train AI systems should be as inclusive as possible, representing different population groups). This also requires adequate respect for potentially vulnerable persons and groups (e.g. such as persons with disabilities, ethnic minorities, children, or other groups at risk of exclusion).

The combination of **human experts** with the **proposed AI methodology**, as general outcome of this project, results in significant ***outcomes***. Another critical factor towards the robustness of the current analysis corresponds to possible future demands evolution or change. Consequently, AI solutions are proposed and designed in a way to adapt to several changes, and to cover future challenging requests and scenarios, with slight modifications on their structure.

During **Phase 1** (Section 1.3.1) of the study, several workshops with stakeholders were implemented to agree on the identified list of Business Use Cases (BUCs). As it is explicitly mentioned on the Framework Contract (No 05/2021 UNDER FRAMEWORK CONTRACT No LISA/2019/OP/01/01/01), the list of the BUCs (on **Chapter 3 - Business Use Cases** Chapter 3 - Business Use Cases) considers and adheres precisely to the predefined timeliness of the assignment along with the provided effort. On this direction, a prioritisation exercise was implemented with the stakeholders during the workshops to conclude in which BUCs are the more significant ones, to be analysed and assessed from the AI perspective. However, many of the deprioritised Business Use Cases, which are not reported in the current deliverable, can be considered in future assessments. To summarize, the list of the identified BUCs for the AI in CRRS project, is formulated as follows:

- **BUC 01** - Identification of Risk for a Specific Group of Travellers
- **BUC 02** - Review and Validate Identified Groups of Travellers of Risk
- **BUC 03** - Analyse Risk Profiles Correlations
- **BUC 04** - Automatic Processing of Textual Information from Various Sources
- **BUC 05** - Ex-Post Assessment Process
- **BUC 06** - Virtual Assistance to Provide Guidance on Certain Activities
- **BUC 07** - Verification process when a hit takes place

Critical factors towards the prioritisation, are: (i) the ethical nature of the BUCs, and (ii) the data-sources availability for design and the subsequent implementation of the BUCs. Ethics and AI are two highly connected principles, while the Commission's guidelines for trustworthy AI provided a helpful and robust framework for the principles that AI solutions should adhere to, and to be subsequently applied towards the design of the current's project BUCs. For this purpose, section 1.5 of the current report, ***Protection of Fundamental Rights in AI***, provides the Consortiums' suggested practices towards the protection of fundamental rights, since it is widely acknowledged their paramount importance to eu-LISA. On the other hand, data availability is the necessary initial step of all data-related, Machine Learning, and AI initiatives. The sufficiency of the amount of data differentiates among the various scenarios, and BUCs; while precise rules which can guarantee how many data we need for the case studies, cannot be concretely define at this stage of the assessment. AI and Machine Learning are two crucial technology fields, able to extract high-quality results if the amount and the structure of input data is sufficient and of high-quality. Consequently, as

a Consortium, an important recommendation towards the success of the current and future initiatives, corresponds to the best possible exploitation of the input data sources, as extracted from the various CBSs. For the current assessment, we consider data extracted from ETIAS, VIS, and EES; while SIS is another data source which is important for future investigation from the AI perspective.

In **Phase 2 (**Section **1.3.2)**, Lot1 proceeded with the analysis of the relevant stakeholder's documentation, while special emphasis was dedicated on the analysis of the Frontex processes concerning the screening rules. Several workshops with the stakeholders were performed to cover several business and technical aspects of the project, and to further understand eu-LISA's Stakeholders expectations. Main actions of Phase 2 included the requirements gathering, along with their elicitation according to the BABOK (Business Analysis Body Of Knowledge) guide. Additionally, the key methodology towards the identification of the specific AI aspects. i.e., the so-called DMADV (i.e., Define, Measure, Analysis, Design, Verify) (**Section 10.5**) is first presented during Phase's 2 activities. Specifically, this methodology composes of five distinct steps; and the first two, corresponding to the "problem definition", and "measure", are described during Phase's 1 and Phase's 2 activities. During "problem definition" step, a high-level assessment of how risk profiling, screening rules and risk indicators are currently being handled is implemented; while the first actions on how AI is able to enrich and optimise the already existing process is settled. "Measure" step is responsible for gathering all necessary requirements that capture the current technological framework within eu-LISA; while identifies interactions among the various CBSs.

In the last phase of the project (**Phase 3 – Section 1.3.3**), the subsequent steps of the DMADV methodology, i.e., "Analysis", "Design", and "Verification" are analysed. Specifically, in the "Analysis" part, Lot1 performed a detailed investigation that revealed all necessary information so that the AI BUCs can produce actionable outcomes. This step relies on the key concepts of *Machine Learning Operations (MLOps)*, and *Data Operations (DataOps)*, while the complete assessment is described on **Section 5.3**. During the "Design" step, the concrete assessment of the AI tools and frameworks is implemented, concluding into the following AI-related applications which are of special interest for eu-LISA: (i) AI for Decision Making; (ii) Clustering Analysis; (iii) Anomaly Detection; (iii) Conversational Agents and Virtual Assistants. Finally, the current methodology concludes with the "Validation" (or verification) step, resulting into the applicable AI technologies (i.e., **Sections 5.3** and **5.4**), and the assessment of the foreseen implementation costs.

Concluding, if the current Study will be promoted into an implementation project, a Proof of Concept (PoC) can be performed in order to illustrate the functionality of the proposed system; and to further highlight the performance, advantages and constraints after eliciting Business and Stakeholder Requirements. Additionally, in terms of the fundamental rights application, PoCs could assess implementations of legal base compliancy.

7 Appendix I - Proposed Technological Frameworks

In this Chapter, we provide a detailed overview of the Consortium's suggested tools and frameworks able to support all building blocks of the proposed DSML platform. To perform the current study, Lot1 took into consideration the current architectural landscape of CRRS, along with the already existing components. Nevertheless, this study relies only on open-source frameworks and tools, and for this purpose, all following tools rely only on open-source technologies. In case the current study becomes a detailed design and implementation project, the Consortium will be able to ensure that the existing CRRS counterparts are going to be considered and further ensure that no duplicate components will be included.

To support the selection of the appropriate tools for the various layers of the proposed Data Science and Machine Learning (DSML) architecture, we perform a thorough comparison for the various building blocks. Demonstrates the proposed services and open-source tools/ technological frameworks, per building block considered for all BUCs, while in the individual sections for the specific Use Cases, we explain in further detail our selections.

Building Block	Service	Open-Source Tool
Data Ingestion & Integration	Ingestion	Apache Kafka, Apache NiFi, Apache Flink, Apache Storm, Apache Flume
	Integration (ELT/ETL)	Apache NiFi, Apache Kafka, Apache Flink, Apache Spark, Apache Storm, Apache Gobblin, Apache AirFlow, Pentaho Kettle, OpenETL
Metadata Management	Data Quality	Hive Metastore, Metadata (Kubeflow)
Data Store	RDBMS, NoSQL	PostgreSQL, SQLite, MariaDB, MySQL, Apache Cassandra, Apache HBase, MongoDB, Neo4j, Apache, CouchDB, OrientDB, Riak, Redis
Data Processing	Orchestrator	Apache Spark, Apache Flink, Apache Beam, Apache Storm
Artificial Intelligence & Machine Learning	AI, ML, and Data Science	• **Tools**: Acumos AI, ClearML, H2O.ai, OpenNN, TensorFlow, PyTorch, Rasa Open Source, spaCy • **Libraries**: TensorFlow, PyTorch, NLTK, Keras.io, SciKit Learn, MLlib, Katib, • **Model Serving Frameworks**: MLFlow, TorchServe, KFServing, Seldon Core, Triton Inference Server • **Supported Coding Languages**: R, Python, Cython, C, C++, Spark, Scala, Java. • **Code Development Notebooks**: JupyterHub, Notebook Server (Kubeflow), Elyra , JupyterLab. • **IDE Development/ Scripting:** R-Studio, NetBeans, Eclipse, Visual Studio Code • **Citizen Data Science Tools:** Weka, Orange, KNIME, MLjar
Data Visualization	Business Intelligence	Pentaho, KnowAge, MetaBase, SuperSet
Self-Service Data Science	Advanced Analytics	MLFlow

Table 20. Proposed Technological Frameworks per Building Block for All BUCs

In this stage of the analysis, we note that *the above table fits the purpose of providing a full-blown approach that will enrich the key objective of the current project*, which is the **AI-enablement of the seven different Business Use Cases** (**BUCs**), as they were explicitly defined within "Chapter 3 - Business Use Cases" of this document, in order to provide robust and flexible solutions that will enhance **eu-LISA's capabilities**, by simultaneously involving strong human intervention.

7.1 Data Ingestion & Integration Tools

Data ingestion can be considered the process of collecting different types of data (i.e., **both batch** and **streaming**) from several heterogeneous data-sources. Data ingestion tools facilitate this process by providing a proper framework that extract data from heterogeneous sources, in a variety of formats, converting, transforming, and modifying the already existing files, and preparing a single larger dataset. The process can be either continuous or asynchronous and can be either real time or batch. The source and sink of files can be of different format or can follow same or different protocols, which may require some form of transformations or conversions. Proper data ingestion tools eliminate the demand for manually coding individual data pipelines for every data source and further accelerate data processing by helping to deliver data efficiently to ETL tools and other types of data integration software or load multi-sourced data directly into a Data Store.

Tools considered in the current analysis for CRRS provide a structure that empowers eu-LISA's data engineers to fetch, ingest, integrate, and perform computations on datasets acquired by disparate data sources. Apart from the data gathering, integrating, and processing perspectives, data ingestion tools facilitate modifying the data for analytics and storage purposes. Additionally, due to the heterogeneous type of data, we include in our assessment both batch and real-time streams. Before proceeding to the selection of the data ingestions tools, we should first be aware of the abovementioned specifications and perform a well-defined comparison among them. On this direction, only few tools are efficient in processing batch data (i.e., data in rest) while others are made only for real-time data streaming (i.e., data in motion).

Criteria	Apache Kafka	Apache NiFi	Apache Flink	Apache Storm	Apache Flume
Latest Stable Release	2.8.1	1.16.0	1.15.1	2.4.0	1.10.0
Prominent Usage	For data ingestion in real-time	For ETL purposes with visualization	For ETL & business purposes	Only for data computation purposes	Ingests high-volume of event-based data into HDFS
Primary written in	Java	Java	Java & Scala	Clojure	Java
License	Open Source	Open Source	Open Source	Open Source	Open Source
Basic Nature	Works well for messaging Streaming data	Works well for data flow creation between different systems	Based on the operator-based computational model	Processes large data volumes & high-velocity data streams. (Over a million records per second per node on medium-sized clusters)	Collects, aggregates, and moves large amounts of log data from many different sources to a centralized data store. Specially designed for Hadoop.
Type of Data	Stream	Batch and Stream Data	Batch and Stream Data	Micro-batch and Stream	Stream

Criteria	Apache Kafka	Apache NiFi	Apache Flink	Apache Storm	Apache Flume
Type of Loading	Event driven	Both (Event and not-event)	Both (Event and not-event)	Both (Event and not-event)	Event driven
Architecture	Process topology	Flow based	Kappa	Master-slave	Deploys as one or more agents. An agent is a JVM process that hosts the components through which events flow from an external source to next destination. Each Agent contains: Source(s), Channel(s) and Sink.
User Interface	Shell command line	Shell command line	Shell command line	Shell command line	Shell command line
Event Prioritization	Programmable	Supported	Supported	Supported	Supported

Table 24 overviews the main data ingestion tools considered for the current analysis.

Criteria	Apache Kafka[66]	Apache NiFi[67]	Apache Flink[68]	Apache Storm[69]	Apache Flume[70]
Latest Stable Release	2.8.1	1.16.0	1.15.1	2.4.0	1.10.0
Prominent Usage	For data ingestion in real-time	For ETL purposes with visualization	For ETL & business purposes	Only for data computation purposes	Ingests high-volume of event-based data into HDFS
Primary written in	Java	Java	Java & Scala	Clojure	Java
License	Open Source	Open Source	Open Source	Open Source	Open Source

[66] https //github com/apache/kafka
[67] https //github com/apache/nifi
[68] https //github com/apache/flink
[69] https //github com/apache/storm
[70] https //github com/apache/flume

Criteria	Apache Kafka[66]	Apache NiFi[67]	Apache Flink[68]	Apache Storm[69]	Apache Flume[70]
Basic Nature	Works well for messaging Streaming data	Works well for data flow creation between different systems	Based on the operator-based computational model	Processes large data volumes & high-velocity data streams. (Over a million records per second per node on medium-sized clusters)	Collects, aggregates, and moves large amounts of log data from many different sources to a centralized data store. Specially designed for Hadoop.
Type of Data	Stream	Batch and Stream Data	Batch and Stream Data	Micro-batch and Stream	Stream
Type of Loading	Event driven	Both (Event and not-event)	Both (Event and not-event)	Both (Event and not-event)	Event driven
Architecture	Process topology	Flow based	Kappa	Master-slave	Deploys as one or more agents. An agent is a JVM process that hosts the components through which events flow from an external source to next destination. Each Agent contains: Source(s), Channel(s) and Sink.
User Interface	Shell command line	Shell command line	Shell command line	Shell command line	Shell command line
Event Prioritization	Programmable	Supported	Supported	Supported	Supported

Table 21. Comparison of the Considered Open-Source Data Ingestion Tools for AI in CRRS

7.2 Data Store Tools

The role of Data Storage during our study is twofold: (i) It defines where the ingested data are coming from; and (ii) it determines the destination of the model results. The two categories of data storage that will be analysed in the context of this study are **Relational Data Stores** and **NoSQL Data Stores**. The differences between the two categories are illustrated on the following Table.

Criteria	Relational Data Stores	NoSQL Data Stores
Definition	Data Stores that use the relational model for data storage and retrieval	Data Stores that use the non-relational model for data storage and retrieval
Structure	Table-based models (i.e., they use rows and columns to hold data in a structured manner)	Underlying structure models
Scalability	Vertically scalable	Horizontally scalable
Data Storage Size	Medium to large amounts of data	Large amounts of data

Table 22. Relational Data Stores vs. NoSQL Data Stores Quantitative Comparison

The main and conceptual difference between the two different approaches is that **Relational Data Stores** have a **fixed schema**, while **NoSQL** data stores can handle **unstructured** or **semi-structured** data. Relational databases are being used traditionally for many years and use SQL (Structured Query Language) to manage data, which makes them simple to use. On the other hand, NoSQL data stores are more flexible solutions and provide great speed regarding the response time in high volume data. NoSQL data stores include different models in accessing and managing data and the most common are the following:

Document Store

Data is stored in document format, as the name suggests. This type of storage can manage effectively semi-structured data, traditionally stored in formats such as JSON, XML or BSON. Their benefits are that they are simple to use and flexible since data schemas do not need to match in the documents. However, this may be considered a problem when it comes to complex transactions.

Key-Value Store

It is considered one of the most common NoSQL data stores due to their simple format. The way data is organized in this schema is a dictionary of key-value pairs, where each item has both key and value. Each key is unique and can only take string values, whilst the value may take other types of data formats, such as String, JSON, XML etc.

Wide-Column Store

The specific storage type, stores data as sections of columns of data instead of rows of data. It combines the characteristics of the relational databases with the key-value stores approach. It is even considered as a multi-dimensional key value store that is persistent distributed, sorted as well as parse.

Graph Store

Designed specifically to store graph-oriented data structures. A graph data structure is any storage system that provides index-free adjacency, i.e., every node contains a direct pointer to its adjacent element without any index lookups. Data elements are stored as nodes, edges and properties. An edge depicts the relationship between nodes.

In the following two tables, the different **Relational** and **NoSQL** data stores that were considered during our analysis are going to be presented respectively:

Criteria	PostgreSQL	SQLite	MariaDB	MySQL
Latest Stable Release	14.4	3.39.0	10.8.3	8.0.29
Programming Language	C	C	C, C++, Perl, Bash	C, C++
License	Open-Source	Open-Source	Open-Source	Open-Source

Table 23. Comparison among the considered Relational Databases

Criteria	Apache Cassandra	Apache HBase	MongoDB	Neo4j	Apache CouchDB	OrientDB	Riak	Redis
Latest Stable Release	4.0.4	2.3.4	5.0.9	4.4.8	3.2.2	3.1.9	3.0.7	7.0.2
Programming Language	Java	Java	C++, JavaScript, Python	Java	C++, JavaScript, ErLang	Java	ErLang	C
License	Open-Source	Open-Source	Open-Source	Open-Source	Open-Source	Open-Source	Open-Source	Open-Source
Data Model Types	Wide-column store	Wide-column store	Document	Graph	Document	Multi-model (supports graphs, document, key/value and object models)	Key-value	Key-value

Table 24. Comparison among the considered NoSQL Databases

7.3 Metadata Management Tools

The following Table overviews the main metadata management tools that are considered for the proposed **DSML architecture**, and further provides their main functionalities. There were two metadata tools considered: Apache Hive Metastore and Metadata Kubeflow. Both tools follow different approaches. **Apache Hive Metastore** is convenient for *simpler metadata scenarios and includes storage of metadata regarding data store for Hive tables and relationships, i.e., schemas and locations, and partitions in relational databases*. Unfortunately, Hive supports only five backend databases: Derby, MySQL, My SQLServer, Oracle and PostgreSQL. On the other hand, **Metadata Kubeflow** is focused only on *managing and tracking machine learning workflows. It is a tool that is focused on Machine Learning applications and not the ecosystem itself.* The main difference between the two tools is the direction of the design of the metadata management.

Criteria	Apache Hive Metastore	(Metadata) Kubeflow
Latest Stable Release	3.1.2	1.5
License	Open Source	Open Source
Description	Relational database that stores metadata related to the tables/schemas to easily query big data stored in HDFS. When a new Hive table is created, the information related to the schema is stored in the Hive Metastore relational database. Other information such as input/output formats, partitions, HDFS locations are all stored in the Metastore.	Tracks and manages metadata of Machine Learning workflows in Kubeflow. In this context, metadata means information about executions (runs), models, datasets, and other artifacts. Artifacts are the files and objects that form the inputs and outputs of the components in your ML workflow.
Key Features	The Hive Metastore provides two important features of a data warehouse: • data abstraction; • data discovery Without the data abstractions provided in Hive, a user must provide information about data formats, extractors and loaders along with the query.	Kubeflow pipelines for metadata management of the Machine Learning process, from data preparation to model training & serving.
Metadata types & instances	Includes the following Hive metadata about the created tables: • A table definition; • Column names; • Data types; • Comments in a central schema repository.	**Model Training Metadata:** • Metrics & Hyper-parameters • Learning curves • Code & config. files • Predictions • Diagnostic charts • Console logs • Hardware logs **Trained Model Metadata:** • Model binary or location • Dataset versions • Links model training runs • Who trained the model • Model descriptions **Artifact Metadata:**

Criteria	Apache Hive Metastore	(Metadata) Kubeflow
		• Paths to dataset/ model • Dataset hash • Dataset/prediction preview • Feature column names (for tabular data) • Who created/modified • Last modified • Size of the dataset

Table 25. Metadata Management Tools considered to serve the DSML Architecture

7.4 Data Processing Tools

The following Table illustrates the considered **Open-Source** tools for the current assessment, applying for both batch and steam processing at massive scale. Stream processing processes the data as soon as it arrives in the store layer, i.e., which would be often to real-time or near-real-time conditions. This would typically be in sub-second timeframes, for the end-user the processing to be implemented in almost real-time conditions. Contradictory, batch processing, includes latencies between the time data appears in the store layer and the time is available on the subsequent analysis layers (i.e., AI & ML layer of the proposed DSML platform). However, in the scenarios where historical data is required for the ML/AI model training, batch processing is a crucial data processing step, especially for instance on time-series analysis problems, in which look-back window frames are key prerequisites. Moreover, a processing category that should be also mentioned, is the Micro-batch Processing, standing as an intermediate ground among streaming and batch processing. It balances latency and throughput, and thus is ideal for several scenarios.

The tools provided on the Table below provide fault-tolerance and data-distribution for heavy and distributed computations, which totally applies on each different eu-LISA's Business Use Cases (BUCs). The prioritisation among the sufficiently big number of different open-source data processing tools is implemented according to their efficiency, flexibility, and high-exploitation (i.e., popularity) by several organisations in the industry and academia domains.

Tool	Apache Spark[71]	Apache Flink[72]	Apache Beam[73]	Apache Storm[74]
Licence	Open-Source (Apache License 2.0)	Open-Source (Apache License 2.0)	Open-Source (Apache License 2.0)	Open-Source (Apache License 2.0)
Computation Model	Spark is based on the micro-batch model	Flink is based on the operator-based computational model.	Beam is based on a unified model for defining and executing data-parallel processing pipelines	Distributed & real-time computation system.

[71] https //github com/apache/spark
[72] https //github com/apache/flink
[73] https //github com/apache/beam
[74] https //github com/apache/storm

Tool	Apache Spark[71]	Apache Flink[72]	Apache Beam[73]	Apache Storm[74]
Streaming Engine	Uses micro-batches for all workloads.	Uses streams for all workloads: streaming, SQL, micro-batch and batch. Batch is a finite set of streamed data.	Uses a single API for both batch and streaming data types.	Processes unbounded data streams in real-time with low latency. Storm uses tuples, spouts, & bolts that construct its stream processing topology.
Throughput	High	High	High	Low
Latency	Apache Spark has higher latency as compared to Apache Flink.	Apache Flink's data streaming run-time achieves low latency and high throughput.	Batch & streaming are merely two points on a continuum of latency; for instance, if a batch pipeline is first constructed but the latency requirements change, it can be easily altered to streaming within the same API.	Low latency stream processing by pipelined data transfers, like Apache Flink.
Performance	It is considered the most matured data processing tool in the community. However, its stream processing is not much more efficient than Apache Flink's (which uses micro-batch processing).	Apache Flink uses native closed loop iterations operators which makes ML & graph processing faster.	While it offers great convenience in writing data-parallel processing pipelines, its performance is slower compared to Apache Flink's & Apache Spark's.	Apache Storm is like Apache Flink, in the sense that is real time and event based.
Fault Tolerance	Spark Streaming recovers lost work and delivers exactly once semantics out of the box with no extra code or configuration.	The fault tolerance mechanism followed by Apache Flink is based on Chandy-Lamport distributed snapshots. The mechanism is lightweight, which results in maintaining high throughput rates & provides consistency guarantees.	Apache Beam is fault tolerant.	Apache Storm is fault tolerant. When workers die, will automatically restart them. If a node dies, the worker will be restarted on another node.
Memory Management	Automatic memory management	Automatic memory management	Automatic memory management	Automatic memory management

Table 26. Comparison of the considered Data Processing Tools

7.5 Open-Source AI/ ML Tools

The following table highlights the main **Open-Source** AI platforms, providing a key summary of their functionality, along with their main features. Additionally, we should mention that all considered platforms are active GitHub projects on the time-period where this deliverable is written for the AI in CRRS project.

Tool	Overview	Key Features	Active GitHub Project
Acumos AI **https://www.acumos.org/**	Acumos AI is an open-source platform that makes it easy to build, share, and deploy AI apps. It is a part of the LF AI Foundation, an umbrella organization within The Linux Foundation that supports and sustains open-source innovation in AI, ML and Deep Learning while striving to make these critical new technologies available to developers and data scientists everywhere.	• Acumos marketplace to discover and deploy various AI libraries. • On-boarding support to enable interoperability. • A graphical tool to manage AI models in preparation for a runtime environment. • A community to develop marketplace solutions. • Dockerization support to run AI within a container. • API connectivity and micro-services tools.	Yes
ClearML **https://clear.ml/**	ClearML is an open-source platform that automates and simplifies developing and managing ML-based solutions for thousands of data science teams all over the world. It is designed as an end-to-end MLOps suite that enables users to focus on developing ML code & automation, by simultaneously ensuring their work is reproducible and scalable.	• An ecosystem for experiment management with zero integration hassles. • Experiment orchestration inside containers (development as well as production). • Scheduling of jobs via priority queues and resource allocation. • Remote allocation of computing resources through a single line of command. • The ability to run Bayesian hyper-parameter optimization with zero integration. • Collaborative workspace with optional permission management.	Yes

Tool	Overview	Key Features	Active GitHub Project
H2O.ai https://h2o.ai/	H2O is a fully open source, distributed in-memory machine learning platform with linear scalability. H2O supports the most widely used statistical & ML-based algorithms including Gradient Boosted Machines, Generalized Linear Models, Deep Learning and more. H2O also has an industry leading AutoML functionality that automatically runs through all the algorithms and their hyperparameters to produce a leader board of the best models. The H2O platform is used by over 18,000 organisations globally and is extremely popular in both the R & Python communities.	• H2O.ai contains several leading algorithms, developed from the ground up for distributed computing and for both supervised and unsupervised approaches, including Random Forest, GLM, GBM, XGBoost, GLRM, and Word2Vec, among others. • H2O's AutoML can be used for automating the ML workflow, including automatic training and tuning of many models within a user-specified time limit. • Users can build models exploiting the programming languages they already know, such as R, and Python, among others, to build models in H2O, or use H2O Flow, a graphical notebook based interactive user interface that does not require any coding. • In-memory processing with fast serialization between nodes and clusters to support massive datasets. • Simple deployment with fast and accurate scoring in any environment, also for very large models.	Yes
OpenNN https://www.opennn.net/	OpenNN is an open-source Neural Networks library for ML. It solves many real-world applications in energy, marketing, and health, among others. OpenNN contains sophisticated algorithms and utilities to deal with many artificial intelligence solutions, such as regression, classification, forecasting, and association.	• C++-based software library • Regression analysis to model ML outputs • Data classification to assign specific patterns • Forecasting based on historical datasets • Association mapping between two correlated variables • A neural designer tool to simplify the process of building neural networks	

Tool	Overview	Key Features	Active GitHub Project
TensorFlow	TensorFlow is an end-to-end open-source platform for ML. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that enable researchers to push the state-of-the-art in ML and developers to build easily and deploy ML powered applications. TensorFlow provides easy model building, robust ML production (i.e., training & deployment) in the cloud, on-prem, in the browser, or on device (despite the programming language used by programmers). Additionally, it enables users with the capability of having powerful experimentation for research.	• Support for multiple languages, including JavaScript, which is relatively rare in the open-source AI space. • Intuitive high-level APIs, such as Keras to easily build and train ML models. • Platform-agnostic ML production – on-premises, in the cloud, in your browser, or locally on the device. • TensorFlow Lite for mobile applications and embedded or IoT devices. • Cross-compatibility between AI/ML models that you have trained on different TensorFlow versions. • A wide variety of applications, including predictive analysis, object classification, and conversational AI.	Yes
PyTorch **https://pytorch.org/**	PyTorch is an open source ML framework based on the Torch library, used for applications such as Computer Vision and Natural Language Processing, primarily developed by Meta AI. It is a free and open-source software released under the Modified BSD license. Although the Python interface is more polished and the primary focus of development, PyTorch also contains a C++ interface.	• A production-ready environment powered by TorchServe for quickly deploying models. • A distributed backend architecture to enable distributed training and performance optimization. • Algorithms for Computer Vision as well as Natural Language Processing. • Supported by all major public clouds for flexible development. • End-to-end workflow from Python to iOS/Android for mobile app development. • Native exports possible from Open Neural Network Exchange (ONNX).	Yes

Tool	Overview	Key Features	Active GitHub Project
Rasa Open Source https://rasa.com/	Rasa is an open-source ML framework to automate text-and voice-based conversations. With Rasa, users can build contextual assistants on: Facebook Messenger, Slack, Google Hangouts, Webex Teams, Microsoft Bot Framework, Rocket.Chat, Mattermost, Telegram, Twilio, and other custom conversational channels, or voice assistants, such as: Alexa Skills, and Google Home Actions. Rasa facilitates in building contextual assistants capable of having layered conversations with lots of back-and-forth. For a human to have a meaningful exchange with a contextual assistant, the assistant needs to be able to use context to build on things that were previously discussed. For this purpose, Rasa enables users to build assistants in a scalable way.	• Natural language understanding to convert messages into structured data and analyse intent. • ML-powered dialogue management to drive the assistant's conversation flows based on context. • Built-in integration for 10+ popular messaging channels. • Complete visibility into the AI training pipeline, model design, and underlying code. • Strong community support from 10,000+ forum members. • An optional Rasa X toolset for testing, enhancements, and new updates.	Yes
spaCy https://spacy.io/	spaCy is an open-source software library for advanced Natural Language Processing, written in Python and Cython, published under the MIT license. spaCy provides software for production usage, and further supports Deep Learning workflows by connecting statistical models trained by popular machine learning libraries (i.e., TensorFlow, PyTorch, MXNET) with its own ML library Thinc.	• SpaCy features Convolutional Neural Network models for part-of-speech tagging, dependency parsing, text categorization and named entity recognition (NER). • Prebuilt neural network models perform these tasks, available for 17 languages, including English, Portuguese, Spanish, Russian and Chinese, and there is also a multi-language NER model. • Support for tokenization for more than 65 languages allows users to train custom models on their own datasets as well.	Yes

Table 27. Overview of the main Open-Source AI Platforms

The following table provides a quantitative comparison among the most challenging **AI/ML model training libraries**, used by multiple organisations in both industry, and academia.

Libraries/ Criteria	Latest Stable Release	License	Description	Languages	Supported Algorithms	Use Cases
TensorFlow	2.9.0	Open-Source	End-to-end ML/DL library; Complete Control over building robust DNNs & ML models; Deploy models on web/ cloud/ mobile/ edge; Supports numerous libraries, extensions & tools for integrating responsible AI into ML solutions; GPU/ TPU support	C, C++, Python	Regression, Classification, Clustering, Dimensionality reduction, Data pre-processing, Generative Models, Reinforcement Learning	Time-series analysis, Image/ Video classification/ detection, Audio Processing, Text-based applications
PyTorch	1.12.0	Open-Source	Builds customised DNNs; Production-ready with TorchServe; Distributed computing support; Scalable deployment; Supports wide array of tools extensions to solve complex problems; GPU/ TPU support	C, C++, Python	Regression, Classification, Clustering, Dimensionality reduction, Data pre-processing, Generative Models, Reinforcement Learning	NLP-based applications, Image/ Video Processing, Audio Processing, Time-series analysis,
NLTK	3.7	Open-Source	Supports more than 50 language datasets & trained language models; while it serves as a wrapper for industrial-scale NLP libraries; easily interacts with TensorFlow, PyTorch & Keras.	Python	Text Classification, Segmentation, Clustering, etc.	NLP based applications, Text Classification, segmentation, stemming, tokenisation, tagging, parsing
Keras.io	2.9.0	Open-Source	Advanced open-source Python deep learning API and framework built on top of TensorFlow; more-user friendly (high-level API of TensorFlow); GPU/TPU support	Python	Neural networks API	Time-series analysis, Image/ Video classification/ detection, Audio Processing, Text-based applications

Libraries/ Criteria	Latest Stable Release	License	Description	Languages	Supported Algorithms	Use Cases
SciKit Learn	1.1.1	Open-Source	Machine Library providing efficient tools for statistical modelling	C, C++, Cython, Python	Regression, Classification, Data Pre-processing, Clustering, Model selection, etc.	Statistical analysis, Probabilistic analysis, Time series analysis, Descriptive/ Predictive Analytics
MLlib	Included in Spark's 3.3.0 release	Open-Source	Spark's scalable machine learning library consisting of learning algorithms & utilities, while it provides access to APIs that enable deep learning in few lines of code.	Python, Spark	Classification, Regression, Naive Bayes, Decision Trees, Random Forests, Recommendation Algorithms, Clustering, Frequent Item-sets, Association Rules, etc.	Image/ Video Processing, Audio Processing, Time-series analysis, Statistical analysis, Probabilistic analysis,
Katib	v0.14.0-rc.0	Open-Source	Kubernetes-native project for automated machine learning (AutoML). It supports Hyperparameter Tuning, Early Stopping & Neural Architecture Search; while interacts with TensorFlow, Apache MXNet, PyTorch, XGBoost, etc.	Any Language, Kubeflow framework	Random Search, Grid Search, Bayesian Optimization, Population-based training, Tree of Parzen Estimators (TPE), etc.	AutoML, DNNs, Hyperparameter tuning, Early stopping and Neural Architecture Search (NAS)

Table 28. Open-Source Machine Learning/ Artificial Intelligence Model Training Libraries - Quantitative Comparison

Additionally, the following Table provides a quantitative comparison among the most challenging **AI/ML model serving** libraries, used by multiple organisations in both industry, and academia.

Criteria	Latest Stable Release	License	Description	Key Features
TensorFlow Serving	2.8.2	Open-Source	Flexible, high-performance serving system for ML models, for production environments. TensorFlow Serving makes it easy to deploy new algorithms & experiments, while keeping the same server architecture & APIs. TF Serving provides integration with not only TF models but can be easily extended to serve other types of models & data.	• Serves multiple models, or multiple versions of the same model at the same time. • Exposes both gRPC and HTTP inference endpoints. • Allows deployment of new model versions without changing the source code. • Provides flexibility on experimenting with different models. • Minimal latency to inference time • Supports multiples servables, such as: TF models, Embeddings, Vocabularies. Feature Transformations, & non-TF machine learning models.
TorchServe	0.6.0	Open-Source	Flexible & easy to use tool for serving PyTorch models. It is an open-source framework that makes it easy to deploy trained PyTorch models at scale without having to write custom code. TorchServe delivers lightweight serving with low latency, to deploy ML models for high-performance inference.	• Multi-model serving. • Model versioning for A/B testing. • Metrics for monitoring. • RESTful endpoints for application integration. • Supports any ML environment. • TorchServe can be used for many types of inference in production settings. • Provides an easy-to-use command-line interface.
KFServe	0.6.1	Open-Source	KServe enables serverless inferencing on Kubernetes and provides highly performant, high abstraction interfaces for common machine learning (ML) frameworks such as TensorFlow, XGBoost, scikit-learn, PyTorch, and ONNX to solve production model serving use cases.	• Simple, pluggable, and complete story for production ML inference server by providing prediction, pre-processing, post-processing and explainability. • Customizable Inference Service to add resource requests for CPU, GPU, TPU and memory requests and limits. • Batching individual model inference requests. • Traffic management. • Scale to and from Zero. • Revision management. • Request/Response logging. • Scalable Multi Model Serving.

Criteria	Latest Stable Release	License	Description	Key Features
Seldon Core	v1.14.0	Open-Source	Open-source platform with a framework that makes it easier and faster to deploy your machine learning models and experiments at scale on Kubernetes. It is a cloud-agnostic, secure, reliable and robust system maintained through a consistent security and updates policy.	• Easy way to containerize ML models using pre-packaged inference servers, custom servers, or language wrappers. • Powerful and rich inference graphs made from predictors, transformers, routers, combiners, etc. • Metadata provenance to ensure each model can be traced back to its respective training system, data, and metrics. • Advanced and customizable metrics with integration to Prometheus and Grafana. • Full auditability through model input-output request (logging integration with Elasticsearch).
Triton Inference Server	22.06	Open-Source	Part of the NVIDIA AI Platform, streamlines AI inference by enabling teams to deploy, run, and scale trained AI models from any framework on any GPU- or CPU-based infrastructure. It provides AI researchers & data scientists the freedom to choose the right framework for their projects without impacting production deployment. It also helps developers deliver high-performance inference across cloud, on-prem, edge, and embedded devices.	• Supports multiple DL frameworks (TensorFlow GraphDef, TensorFlow SavedModel, ONNX, & PyTorch). • Simultaneous model execution on the same GPU or on multiple GPUs. • Dynamic batching. • Extensible backends. • Supports model ensemble. • Metrics in Prometheus data format indicating GPU utilization, server throughput, and server latency.

Table 29. Open-Source Machine Learning/ Artificial Intelligence Model Service Libraries - Quantitative Comparison

The following table provides a quantitative comparison among the most well-exploited **interactive notebooks solutions for developing AI/ML code**, as proposed by several organisations in both industry, and academia.

Criteria	Latest Stable Release	License	Description
JupyterHub	2.3.1	Open-Source	Multi-user hub that spawns, manages, and proxy's multiple instances of the single user Jupiter notebook serve. A JupiterHub distribution is tailored towards a particular set of use cases. The two popular use cases are: • Zero to JupyterHub on Kubernetes, for running JupyterHub on top of Kubernetes. This can scale to large number of machines & users. • The Littlest JupyterHub, for an easy to set up & run JupyterHub supporting 1-100 users on a single machine.
Notebook Server (Kubeflow)	Included on Kubeflow's 1.5 release	Open-Source	Kubeflow Notebooks provides a way to run web-based development environments inside your Kubernetes cluster by running them inside Pods, while among the key features included, are the following: • Native support for JupyterLab, RStudio, and Visual Studio Code (code-server). • Users can create notebook containers directly in the cluster, rather than locally on their workstations. • Admins can provide standard notebook images for their organization with required packages pre-installed. • Access control is managed by Kubeflow's RBAC, enabling easier notebook sharing across the organization.
Elyra	v3.10.0	Open-Source	Elyra is a set of AI-centric extensions to JupyterLab Notebooks, while it currently includes the following functionalities: • Visual Pipeline Editor. • Ability to run notebooks, Python or R scripts as batch jobs. • Reusable Code Snippets. • Hybrid runtime support based on Jupyter Enterprise Gateway. • Python and R script editors with local/remote execution capabilities. • Python script navigation using auto-generated Table of Contents • Notebook navigation using auto-generated outlines using Table of Contents. • Language Server Protocol integration. • Version control using Git integration

Criteria	Latest Stable Release	License	Description
JupyterLab	3.0	Open-Source	JupyterLab is a next-generation web-based user interface for Project Jupyter, while it enables users to work with documents & activities such as Jupyter notebooks, text editors, terminals, and custom components in a flexible manner. Key activities include: • Code Consoles which provide transient scratchpads for running code interactively, with full support for rich output. • Kernel-backed documents enable code in any text file (Markdown, Python, R, etc.) to be run interactively in any Jupyter kernel. • Notebook cell outputs can be mirrored into their own tab, side by side with the notebook, enabling simple dashboards with interactive controls backed by a kernel. • Multiple views of documents with different editors or viewers enable live editing of documents reflected in other viewers. JupyterLab works out of the box with JupyterHub 1.0+ and can even run side by side with the classic Notebook.

Table 30. Interactive Notebooks for AI/ML Code Development

The following table provides a quantitative comparison among the well-exploited and **Open-Source citizen data science platforms/ self-service analytics platforms**, depicted as **no/low-code AI solutions**, for both industry and academia. Key motivation behind the investigation of such a solution, lies in the fact that minimum human coding effort is required on these platforms. However, for heavy computational tasks, the AI tools (i.e., including programming languages, libraries, APIs, and frameworks) stand as the best possible candidate for the current eu-LISA's landscape. In future, as the organization continuous to evolve the AI journey, self-service/ citizen data science platforms, may also be considered.

Tool	Licence	Latest Stable Release	Description
Weka[75]	Open-Source (GNU General Public License)	3.9.6	WEKA is a solutions ecosystem engineered to solve Accelerated DataOps challenges, delivering Reference Architectures and Software Development Kits with leading AI solutions partners. **Key Features:** • Production-ready storage solution where the entire data pipeline workflow—ingest data, to batch feature extraction, to hyper-parameter optimization, & finally to inferencing and versioning. • Can be run be executed on the same platform, whether running on-prem or in the public cloud. • Direct access to data for training & inferencing eliminates data staging at the compute layer & storage silos which results in shorter epoch times. • Supports several standard data mining tasks: data pre-processing, clustering, classification, regression, visualization, and feature selection. • Model Deployment: Not explicitly reported on the official documentation, however the integration with other programming languages (R, Python) enables this feature. • Real-time scoring: Not explicitly reported on the official documentation, however the integration with other programming languages (R, Python) enables this feature. • GPU acceleration[76] or via any R, Python integration

[75] https //www cs waikato ac nz/ml/weka/index html
[76] https //github com/Waikato/weka-3 8/blob/master/packages/internal/kerasZoo/src/main/java/weka/classifiers/keras/KerasZooClassifier java

Tool	Licence	Latest Stable Release	Description
Orange[77]	Open-Source (GPLv3 or later)	3.32	Orange is an open-source data visualization, machine learning and data mining toolkit. It features a visual programming front-end for explorative rapid qualitative data analysis and interactive data visualization. To explore data with Orange, one requires no programming or in-depth mathematical knowledge. **Key Features:** • Python API, requires C, C++ compiler (on Win, Microsoft Visual Studio Build Tools are the recommended). • Easy installation; can be installed either with Conda or pip. • Orange includes a wide array of workflow templates designed to help you get familiar with the application. • Building blocks of data analysis workflows that are assembled in Orange's visual programming environment • Add-ons area unit accessible to increase the practicality: it is potential to increase the functionality of Orange through add-ons that area unit accessible on-line. • Responsive GUI • Provides online support • Creates dataset from any graph • Model Deployment: Not explicitly reported on the official documentation; implemented by external modules, upon integration (i.e., for instance Python APIs) • Real-time scoring: Not explicitly reported on the official documentation

[77] https //orangedatamining com/

Tool	Licence	Latest Stable Release	Description
KNIME[78]	Open-Source (GNU General Public License)	4.4.4	KNIME Analytics Platform is the open-source software for creating data science. Intuitive, open, and continuously integrating new developments, KNIME makes data understanding and designing data science workflows and reusable components accessible to all users, without requiring heavy programming knowledge. **Key Features:** • Builds Data Science Workflows: Drag & Drop style graphical interface, without the need for coding. • Visual Programming with nodes for the Entire Data Science Life Cycle. • Tool Blending: including scripting in R & Python, ML, or connectors to Apache Spark. • Blends Data from Any Source: simple text formats (CSV, PDF, XLS, JSON, XML, etc.), unstructured data types (images, networks, documents, etc.), and time-series data; connect to a host of databases and data warehouses. • Scalability through sophisticated data handling (i.e., intelligent automatic caching of data in the background while maximizing throughput performance). • High, simple extensibility via a well-defined API for plugin extensions. • Intuitive user interface. • Import/export of workflows (for exchanging with other KNIME users). • Parallel execution on multi-core systems • Command line version for headless batch executions. • Enables model deployment. • Enables real-time scoring.

[78] https //www knime com/knime-analytics-platform

Tool	Licence	Latest Stable Release	Description

MLjar[79]	Open-Source	0.11.2	• The Open-Source Software (OOS) version of MLjar is Mercury[80], which turns Python notebooks to web applications, and thus non-technical users can process the widgets and execute notebooks with new parameters.
			• For Automated ML: **MLJAR Automated Machine Learning for Humans:** https://github.com/mljar/mljar-supervised
			  ○ Works with tabular data; supervised learning; exploits several different algorithms; provides automatic reporting
			  ○ Model deployment & real-time scoring: Not explicitly reported on the official documentation; implemented by external modules, upon integration (i.e., for instance Python ML libraries)

Table 31. Comparison among Open-Source Citizen Data Science/ Self- Service Platforms

[79] https://github.com/mljar/mljar-supervised
[80] https://mljar.com/mercury/

Finally, the table below provides a comparison among the most prominent **Open-Source** tools for building **robust conversational agents**, i.e., **chatbots**. At their majority, conversational agents exploit Natural Language Processing (NLP)-based and AI formulations in their architectural design. In this project, a chatbot solution is designed to cover the needs of "**BUC.06 – Virtual Assistance to Provide Guidance to certain activities**".

Tool	Licence	Latest Stable Release	Description
Rasa Open Source[81]	Open-Source (Apache-2.0 license)	3.1	Rasa Open Source is an open-source machine learning framework for automated text and voice-based conversations. Understand messages, hold conversations, and connect to messaging channels and APIs. **Key Features:** • **Build Assistants:** ○ Easy Installation via Virtual Environment Setup, or via Python Environment Setup, or build from Source (Github); ○ Integrates dependencies for spaCY and MITIE; ○ Flexible Command Line Interface; ○ Enables conversation-driven development; ○ Generates Natural Language Understanding (NLU) data; ○ Writes Conversation Data, including: stories and rules that formulate the training data for the Rasa assistant; • **Build Assistants - Preparing for production:** ○ **Connecting to a channel:** Rasa Open Source provides many built-in connectors to connect to common messaging and voice channels. Users are also able to connect to their websites or apps with pre-configured REST channels or build their own custom connectors. ○ **Tuning NLU Models:** Rasa Open Source provides a suggested NLU config on initialization of the project, but as the project grows, it is likely that adjustments might be required to the config to suit your training data. ○ **Testing the Assistant:** Rasa Open Source enables validation and test dialogues end-to-end by running through test stories. Users can also test the dialogue management and the message processing (NLU) separately. ○ **Setting up CI/CD:** Setting up a Continuous Integration (CI) and Continuous Deployment (CD) pipeline ensures that incremental updates to the virtual assistant is improving it, and not harming it. • **Deploy Assistants: (i) Deploy Rasa Open Source**; (ii) **Deploy Action Server (**i.e., A Rasa action server runs custom actions for a Rasa Open-Source conversational assistant – e.g., Rasa SDK Python); • **Training Data Format: (i)** Rasa Open Source uses YAML as a unified way to manage all training data, including NLU data, stories and rules. **(ii)** NLU training data stores structured information about user messages. • **APIs:** (i) HTTP API; (ii) NLU-Only Server

[81] https //rasa com/open-source/

Tool	Licence	Latest Stable Release	Description
Botpress[82]	Open-Source (AGPL-3.0 license)	12.28.1	Botpress is the standard developer stack to build, run and improve Conversational-AI applications. Powered by natural language understanding, a messaging API and a fully featured studio, Botpress allows developers around the globe to build remarkable chatbots without compromise. **Key Features:** • Built-in Natural Language Processing tasks such as intent classification, spell checking, entity extraction, and many others; • A visual Conversation Studio to design multi-turn conversations and workflows; • An emulator & a debugger to simulate conversations and debug your chatbot; • Support for popular messaging channels like Slack, Telegram, Microsoft Teams, Facebook Messenger & an embeddable web chat; • SDK & core editor to extend the capabilities; • Post-deployment tools like analytics;
Wit.ai[83]	Open-Source		The core functionality of Wit.ai is based on two major ML technologies, i.e., Natural Language Processing (NLP) and Natural Language Understanding (NLU). While NLP enables Wit.ai to break customer queries into actionable "entities", NLU extracts meaning out of these entities. **Key Features:** • Enables developers to add a modern natural language interface to their app or device with minimal effort. • Converts sentences into structured information. • Developers do not need to develop NLP algorithms, configuration data, performance and tuning. • Encapsulates the terms of AI/ML and enable users to focus on the core features of their apps and devices. • Provides Voice-enabled Android and iOS apps • Raspberry Pi based home automation commanded by speech • Google Glass apps accepting voice commands

[82] https //botpress com/
[83] https //wit ai/

Tool	Licence	Latest Stable Release	Description
OpenDialog[84]	Open-Source (Apache-2.0 license)		OpenDialog is an open-source conversation management platform that enables organisations to deliver large scale conversational applications - through a solution that they can entirely host and control. It allows flexible integration with different NLP platforms, a powerful conversational model, rich analytics and flexible editorial control. **Key Features:** • Supports both voice and text interfaces • Built for creating bespoke, scalable conversational applications • Solves the problems that off the shelf platforms cannot • API-first approach • Open source - Apache 2.0 license • Own your platform and your data • Map your conversations and your knowledge

Table 32. Comparison of Open-Source Frameworks for Conversational Agents/ Chatbots

7.6 Data Visualization Tools

The table below provides a quantitative comparison among the well-exploited and open-source data visualization tools, exploited by several organisations. The comparison was conducted as part of this analysis regarding the **AI opportunities** in the **eu-LISA's** organization and how they can enhance human interaction on the defined certain **Business Use Cases (BUCs)**. More explicitly, since the Machine Learning/ AI analysis part has been completed, the results are usually shared in Visualization tools where authorized users (i.e. **eu-LISA's** stakeholders - in the current study) have access. The reasons for including a visualization component in the designed architecture are twofold: (i) it is proved to be the most user-friendly approach, enabling the users to comprehend the results easier (also for self-exploratory analysis); and (ii) in the future, **eu-LISA** may agree with certain stakeholders (and generally technical audience) to present them the results of the discussed analysis in a certain visual way that they are going to exploit.

[84] https://github.com/opendialogai/opendialog

Tool	Licence	Latest Stable Release	Description
Pentaho[85]	Open-Source	4.8.0	Pentaho offers Pentaho Business Analytics which is a suite of open-source Business Intelligence applications. It supports data integration and with the help of Online Analytical Processing (OLAP) services and interactive dashboards, helps users visualize their data. **Key Features:** • Intuitive platform, where IT & Business Analysts can access and visualize data easily. • Easy access to data from diverse sources ranging from Excel to Hadoop. • Reporting is fast due to in-memory caching techniques, while the output can be generated in various formats. • Detailed visualisation and easy to understand infographics, with drilling and filters available. Seamless integration with third party applications, such as Google Maps. • The devices supported covers almost every platform: Android, iPhone, iPad, Mac, Web-based, Windows.
KnowAge[86]	Open-Source	7.2	KnowAge is the open-source analytics and business intelligence suite that enables users to combine traditional data and big/cloud data sources into valuable and meaningful information. **Key Features:** • Data federation, mash-up, data/text mining and advanced data visualization • Provide comprehensive support to rich and multi-source data analysis. • The suite is composed of two main modules and four additional plugins that can be combined to ensure full coverage of user' requirements.

[85] https //github com/pentaho
[86] https //github com/KnowageLabs/Knowage-Server

Tool	Licence	Latest Stable Release	Description

MetaBase[87]	Open-Source	0.43	Metabase is a simple & powerful analytics/BI tool which lets anyone within an organization to learn and make decisions from their data, without requiring technical knowledge.
			Key Features:
			• In MetaBase an answer to a question can be visualized in several ways: Numbers, Trends, Progress bar, Table, Pivot table, Line chart, Bar chart, Combo chart, Waterfall chart, Row chart, Area chart, Scatterplot or bubble chart, Pie/donut chart, Funnel, Map.
			• Interactive Dashboards
			  ○ Publish & build in minutes
			  ○ Filters, Cross-filtering enable advanced interactivity, connecting multiple dashboards to build immersive data reports.
			• Self-service Analytics
			• Embedded Analytics

Apache SuperSet[88]	Open-Source	1.5.1	Superset is a modern data exploration and data visualization platform, able to replace or augment proprietary BI tools for many teams, while provides integration with a variety of data sources. **Key Features:** • No-code interface for building charts quickly. • Powerful, web-based SQL Editor for advanced querying. • Lightweight semantic layer for quickly defining custom dimensions & metrics. • Out of the box support for nearly any SQL database or data engine. • Wide array of visualizations to showcase data. • Lightweight, configurable caching layer to help ease database load. • Highly extensible security roles and authentication options • API for programmatic customization. • Cloud-native architecture designed from the ground up for scale.

Table 33. Comparison of the considered Data Visualization Tools

[87] https //www metabase com/
[88] https //github com/apache/superset

8 Appendix II – Assessment of the Foreseen Implementation Effort

8.1 Overview

Considering the communicated needs as revealed from this study, **eu-LISA** has defined the demand to extend the existing data insights landscape, adopting an AI applicable toolset enabled through appropriate frameworks. This results in multiple challenges, as the new scientific data sources needs to be incorporated within the existing integrated data ecosystem and provide information capabilities to surrounding environment entities without the need to address severe changes **Lot1** brings together experience and the outline of the applicable professional experts to address the needs of **eu-LISA's** request.

To this extend, the study has been structured to address all requested requirements as depicted from the Inception and Interim report context. **Lot1** study provides technological adaptations, methodologies and principles tackling the upcoming business and IT challenges of introducing new scientific **AI methodologies** and **Tools** consisting of:

BUC1- Identification of Risk for a Specific Group of Travellers: In the specific BUC, during the definition of the risk indicators, the proposed **AI Solution** is designed to identify patterns from the analysis of historical data available in the CRRS pointing to security, illegal immigration or high epidemic risk for a specific group of travellers (Clusters). *For this purpose, classification approaches perform the identification of the groups of travellers corresponding to the specific groups of risk*.

BUC2 - Review and Validate Identified Groups of Travellers of Risk: In this Use Case we design a robust **AI Solution** that seeks for deviations from the already identified groups of travellers of risk for ETIAS. This BUC is handled as a characteristic scenario of *Machine Learning (ML)-based Anomaly Detection*.

BUC3 - Analyse Risk Profiles Correlations: In this BUC, correlations on the abnormal rates of overstaying/ refusals of entry, for specific groups of travellers among ETIAS and EES will be extracted and will further demonstrate the interconnectivity of the different systems. Additionally, as subsequent step of the analysis, we may consider the correlations among epidemiological reports provided by ECDC and WHO, and ETIAS/ EES systems.

BUC4 - Automatic Processing of Textual Information from Various Sources: In this BUC, an AI Solution identifies patterns or a set of common characteristics from the analysis of text data allowing the identification of a group of travellers of risk (Clusters) that meets the criteria for the validation of a proposed risk profile. Specifically, in this *BUC near real-time processing and understanding of the textual information will be considered.* On this direction, novel *ML/AI formulations are proposed to further categorise documents (i.e., reports) from the various CBSs that reach CRRS*.

BUC5 - Ex-Post Assessment Process: In this BUC, the **AI Solution** enhances the ex-post assessment process by analysing and detecting deviations and proposing the review of the risk indicators. This process can be triggered manually by a user or set within a pre-scheduled timeframe (e.g., 6-months window or smaller time-intervals). *This BUC is also handled as a scenario of Anomaly Detection*.

BUC6 - Virtual Assistance to Provide Guidance on Certain Activities: In this BUC, a **virtual assistant AI Solution** (enriched by concrete human intervention) will be implemented to improve several critical business processes, and further extract suggestions on how to contact customer support teams and to further report incidents. *Chatbot solutions, Natural Language Processing, and Deep Learning formulations are considered the best possible candidates which generate concrete answers to certain questions.*

BUC7 - Verification process when a hit takes place: In this BUC, we propose a robust AI Solution during the verification process when a hit takes place. *The specific BUC is also handled as a characteristic scenario of Machine Learning (ML)-based Anomaly Detection (i.e., like BUC.02 and BUC.07)*.

8.2 Assumptions on the foreseen Effort and Planning

The plan has been based on the assumption, that a project of ***-12- calendar months duration*** incorporating all the Business Use cases, shall take place as per the confirmation of the latest received confirmation of the study deliverables. The proposed plan is analogous with the task deliverables, activities, tasks and phases as depicted on the study established phases. Our approach has adapted all necessary information presented on this study and has tailored an indicative plan considering the following also assumptions:

- The Main project is divided into **seven (7) parallel streams of work** to support the establishment of the identified BUCs.
- For every parallel stream, representing a separate use case, we foresee a Lean start-up process at first to produce an **MVP**. This is a crucial milestone for the achievement of each use case as it closely couples the progress of the project and
- All development tasks include a series of Sprints running in Agile, with **2-weeks duration** each.
- The level of parallelism proposed in the foreseen plan is based on the current information of this study and identified activities and therefore can be adjusted accordingly based on the Initiation phase (business justification, project scope & organization) and Planning phase exchanged information with respective EU-Lisa's stakeholders.
- The complete duration of the anticipated effort is **-12- calendar months**.
- No external delays are considered, in this indicative duration from third party services or owners.

8.3 Methodology and Work Packages Proposed

The potential project strategy for the depicted use cases, is proposed to be based upon ***PM2*** framework, whereas the generic implementation and related ones AI may follow the ***Agile*** and ***CRISP-DM*** methodology. A high-level description of the plan phases will follow, which clearly articulates and leverages proposed methodologies for a successful establishment and adoption considering the AI related Business Use Cases (BUCs) within EU-LISA's ecosystem. Lot1 defines the potential project running ***into Phases, Work Packages and Deliverables*** under a combination of **PM2**[89] and **CRISP-DM** processes. Every phase includes apart from the technical deliverables, several deliverables addressing business and technical aspects of the activities performed and actual outcomes. Within this proposal, it is deemed necessary to describe the anticipated roles required to support a potential project, along with necessary assumptions, with a high-level estimate delivering the prescribed BUCs.

[89] PM2, Home | PM² Project Management Methodology (europa eu)

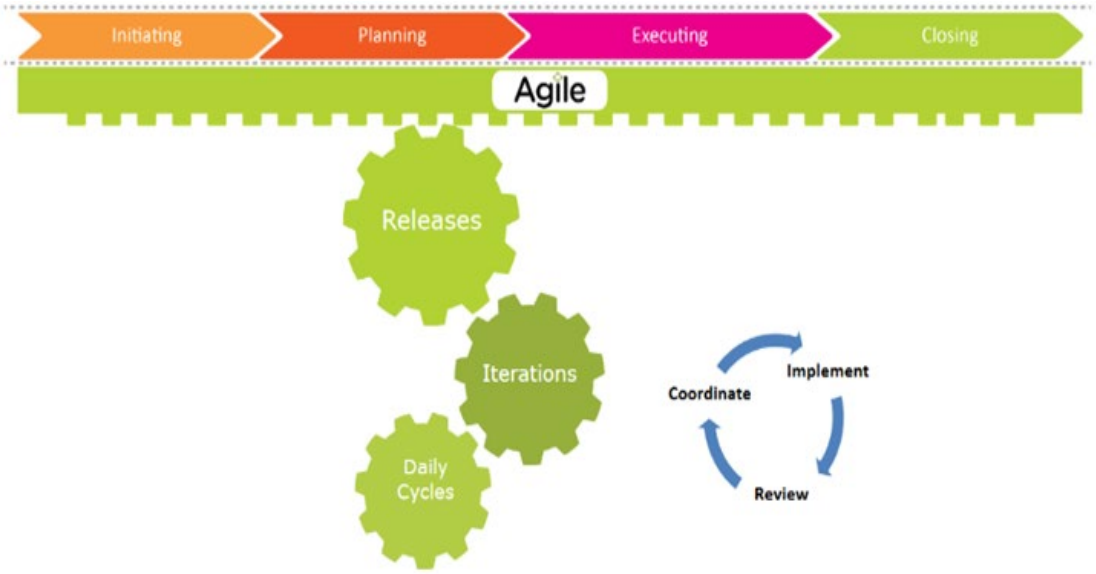


Figure 37. PM2 Methodology, Initiating, Planning, Executing and Closing

The proposed framework demonstrated under this section considers the BUCs establishment after a potential project duration of ***12 calendar months***. In the schedule to be demonstrated, we tried to combine the Agile development approach with the PM2, which is the official *Project Management Methodology* of the *European Commission* and will include the respective four steps according to requested Business Use Cases (01 to 07), as imposed by this methodology *(**Initiating, Planning, Executing,** and **Closing**)*. Therefore, our approach has tailored this methodology and adapted accordingly to meet **eu-LISA's** requirements. This combination had produced a slide modified Gantt that is fully depicted in following sub-sections along with respective assumptions and remarks. Artefacts are delivered under PM authority in all phases. Thus, the whole Gantt is structured under the Initiating, Planning, Executing and Closing phases. Already depicted tasks by the proposed schedule to deliver the AI tools based on the identified business use cases, have been included mostly under the Executing phase.

The following activities depicts in high-level the proposed phases of the project and is being separated into indicative, though tangible deliverables. Respective analysis, implementation and delivery activities have been fully incorporated under the Execution phase of this framework. Sub-sections to follow shall describe in detail the foreseen activities and deliverables per Project phase.

Contract Initiation Phase

WPA Initiating The initiation of the project is an obligatory activity and merely involves the **eu-LISA** and the lead resources of the foreseen project. The proposed tasks include, but not limited to:

WPA.1 Kick-off meeting execution
Participation to Kick-of meeting for the contract, onsite or offsite and preparation of the draft agenda for this meeting.
WPA.2 Understand the actions to perform / Contract
Activity Understanding of the content, context and organization of the contract.
WPA.3 Business Use Cases definition activity
Specify the Business Use Cases (BUCs) that are included in the scope and need to be delivered.
WPA.4 Identify Risks Activity
Identify along with eu-LISA's stakeholders' potential risks and generate a mitigation plan.
WPA.5 Conduct Meetings Activity
Conduct all required meetings.
WPA.6 Identify initial technological strategy Activity (i.e. Open Source)
Identify and describe contract-level technology strategy.
WPA.7 Operational Model Overview as Deliverable
The operational model overview that will govern the project.

Contract Planning Phase

WPB Planning: The project members upon certain request, shall perform a scoping refinement and set the important strategic factors that will drive the preparations, cutover and transition period of the project delivering the AI related BUCs. However, the anticipated Project plan and respective agenda shall be established based on the High-level analysis of the decision on task lists identified under mutual agreement with eu-Lisa's stakeholders, prior. Consequently, this will lead to derive the exact Project estimates, and the estimate on resources needs. This phase includes Project Planning artefacts such as Constraints, Dependencies, Assumptions, Exclusions, scope, deliverables and project timeline.

WPB.1 Tailor Agile process as to match the need of the organization.
WPB.2 Determine Business Objectives for each defined activity and especially on the selected Bucs.
WPB.3 Assign Roles & Responsibilities forming the A-PCT team.
WPB.4 Prepare Overall Schedule as part of the PM artefacts.
WPB.5 Overall Plan for testing and deployment as part of the PM and A-PCT artefacts.
WPB.6 Scoping Refinement (i.e. of the MVPs) as an obligatory activity between the **eu-LISA** stakeholders and project team.

Contract Execution Phase

WPC Analysis: A thorough research is imperative before the entire process initializes, as it allows the assigned project team to understand the current system status, define deltas comparing to the initial study, **eu-LISA's** exact architecture affecting the project, and future planning. **Lot1** main actions in this phase involves thorough analysis of the selected new data sources and sections, eligible for integration under the existing eu-LISA's environment. We foresee this to include business research, data source analysis, system architecture, and application research.

The deliverables include:

(WPC.1) A business process mapping analysis (*CRISP-DM Business Understanding*) affecting the existing platforms as defined through this study per Business Use Case.
(WPC.2) A GAP Analysis state denoting the difference (gaps) between the actual "**As-Is**" state, the defined state as depicted by this study and the "**To-Be**" state of the AI Business Use Cases (BUC) achievement, identifying any overlapping or duplicated activities. This analysis will determine what should be modified in the existing eu-LISA's data landscape and generate planning efforts applicable for any Time plan changes.
(WPC.3) A System Architecture identification Analysis which corresponds to a comprehensive research and analysis on eu-LISA's Data Platform's architecture deployments along with a thorough analysis on the building blocks (*i.e. Platforms, Access Layers, Reporting, SW, HW and network landscape*), system scalability, system data flow, analysis of the surrounding applications environment *(i.e. Data integration etc.*) and other aspects (*i.e. Servers processing, storage devices Input/output operations, and network bandwidth*).
(WPC.4) Data Source Analysis (i.e., *CRISP-DM, Data Understanding*)**,** witnesses a collection of information related to the new Data source and data sections (*i.e. early data profiling, AI Business Use Cases exploration, data quality verification*).
WPC.4.1: This deliverable depicts the business context of the data source, what are the main entities, what are the relationships between each other and in general understand the structure of the information that is offered and is intended to be the source of information supporting the BUCs.
WPC.4.2: The format, the type and volatility of the data stored in the source *(flat files, unstructured and semi-structured, etc.)*.
WPC.4.3: Source data profiling to understand data source capacity, its volumes, structure, integrity and consistency.
WPC.4.4: The deployment structure and data security policy and how it is affected from the indicative list of the Open-Source software selected to support the enablement of the **Business Use Cases (BUCs)**.
WPC.4.5: The integration information including the existing and advanced features at the data source access layer, such as technical interfaces that are offered by the data source for data extraction and delta extraction capabilities of the source.
Additional deliverables of this WP may consider detailed reports, tasks and research documentation that are generated as outcome for each of the above activities**.**

WPD DSML Platform Realization: A proposed blended team by Lot1, corresponding and addressing analysis deliverables is considered at this phase. This phase is separated also to activities required to deliver a design-wise establishment of the building blocks and environments that the common AI use cases shall be operating.

WPD.1 Logical Architecture. Considers the Logical modelling and the interfaces that should be exposed by the Business Use Cases (BUCs) establishment to consuming layers. This work-package depicts the lined of work applicable to be made for all -7- BUCs that need to be reflected on the solution's Logical architecture which:
WPD.1.1 Articulates the data extraction mechanisms and the ways the technical interfaces of the source will be used.
WPD.1.2 Describes the technical interfaces through which data from the DSML design and moreover from the AI processes will be exposed to its users.
WPD.1.3 Includes the early steps of CRISP-DM AI Data Preparation process and Data oriented process required on the DSML platform layer.
WPD.1.3.1 Refers to the data activities on early selecting, cleaning, constructing, formatting and evaluate data model applicable for the early validation of the relational structure required in the AI process and furthermore on the hosted DSML platform.
WPD.1.3.2 Refers to activities of selecting, generating, building and assessing related relational models under CRISP-DM process of respective data and metadata information of the **AI BUCs**. An early evaluation of the models is addressed at this stage, which will further be adjusted on the establishment phase. Particularly, the Design of applicable Data Models, including the definition of Data Area type (persistent and/or not persistent), the requirements complexity and the underlying business processes, the advanced type of Data objects such as Address complexity and performance issues *(i.e. point-in-time (PIT) tables, bridge tables, reference tables)*. The Business Data definition, incorporating the way to apply the soft and hard business rules that have been defined during early steps of this phase. The Metadata Data model which captures and processes the following elements: The business metadata (*the meaning of data for the business*), the technical metadata (*the technical aspects of data, including data types, lineage, results from AI data profiling*), the process execution metadata (*i.e. statistics about running AI processes, including the number of records AI cases sourced and loaded to the destination systems*) and the error tracks which is the primary location to store error information (*i.e. records rejected by the AI quality or inference process and stored in separate Data areas characterized as rejected records*).
WPD.1.4 refers to the modules of the proposed DSML platform incorporating and hosting the selected BUCs, which will be used and the reasoning behind this, addressing eu-LISA's existing environment's policy requirements as security, configurability, extensibility, and performance is addressed. Deliverable of this activity is considered the solution's Logical Architecture along with respective prerequisites as mentioned prior.
WPD.2 Capacity Planning & Physical Architecture. This activity considers the proposed technological platform, the AI BUCs authorized user's availability, the performance requirements and the Data Source Analysis report and mainly the information it provides about the data volumes of the source they prepare:
WPD.2.1, the capacity plan to define infrastructure needs in terms of capacity and processing power.
WPD.2.2, the Physical Architecture design that describes what exactly is needed in terms of infrastructure and how this infrastructure will be interconnected and configured to realize the described logical architecture. Deliverable of ***WPD.2*** activity is considered the solution's *Physical Architecture.*
WPE Data Platform Establishment: This phase depicts the necessary actions and activities of implementing, integrating, testing and deploying the complete solution of the selected BUCs under the approach of a DSML platform.

- ***WPE.1 Infrastructure installation & configuration*** corresponds to the installation of the environments (*PROD,UAT/QA,DEV*) under EU-LISA's landscape considering necessary Containers, VMs, OS, Storage, Network, security policies users & domains, configuration and integration, respectively.
- ***WPE.2 SW installation & configuration*** corresponds to applicable software open-source elements that shall be installed and configured for generating the AI application layer and integrations with identified consuming layers.

WPF AI BuC MVPs: This phase is the cornerstone of the project. The concept is to proceed at first with a Lean start-up producing **7 MVPs** as part of the **BUCs context**. For the completion of the **MVP** series **-6- sprints** are anticipated. The 1st sprint among the proposed pairs may include the analysis and design of the BuC aspects, whereas the 2nd sprint includes the implementation and realization amongst the project team and eu-LISA of the produced results. *It is suggested that sprint of selected BuCs (selection may differ from the order depicted in this study) 1 to 4 to run in parallel for quicker results*. Upon completion of the MVPs the project is proceeding with the necessary

implementation of the rest of the work anticipated to achieve the depicted Business Use cases.

- Sprint x BUC 1 &2 MVP
  - BUC 1&2 MVP analysis and design
- Sprint x+1 BUC 1 &2 MVP
  - BUC 1&2 MVP implementation and realization of the results
- Sprint x BUC 3 &4 MVP
  - BUC 3&4 MVP analysis and design
- Sprint x+1 BUC 3 &4 MVP
  - BUC 3&4 MVP implementation and realization of the results
- Sprint x+2 BUC 5,6 &7 MVP
  - BUC 5,6 &7 MVP analysis and design
- Sprint x+3 BUC 5,6 &7 MVP
  - BUC 5,6 &7 MVP implementation and realization of the results
- Sprint x+3
  - Realization and acknowledgement process among stakeholders

WPG AI BuC establishment:

- ***WPG.1 AI Models,*** depicts the finalization of the AI mechanisms to employ data structure and provide relations data and metadata information as data sets ready to be consumed. Under this step a CRISP-DM re-evaluation of the existing implementation and an early deployment of the produced data structures are considered. Implemented AI/ML models (*considered as deliverables of this task*) are integrated as per eu-LISA's approach to the DSML platform, leveraging access from data landing zones (*i.e. Staging zones*) for performing the extraction of data and metadata, applicable for the Layers of Business systems or Business consumers.
- ***WPG.2 Data Modelling.*** This activity generates the Data Model design, as part of the AI BUCs output, which describes the ways that data is structured and securely made available to the Business consumers. Project experts should differentiate the process when Batch or Real time latency is required.
- ***WPG.3 Data Consumers interface establishment:*** Delivers the AI interfaces integration with other platforms. Apart from necessary integration tasks, a specification is generated as the artefact that will be used by its consumers as input to implement and put in place the required mechanisms to consume data and send it for Decision making under a human intervention approach (No prescriptive approach is depicted as per the conducted study). This specification is the contract between the AI deliverables and its consumers, such as all non-functional requirements that the interface should fulfil regarding results, secure access, availability and performance.
- ***WPG.4 Data Pipelines:*** Through this activity, ETL/ELT pipelines shall be developed for establishing the data sets applicable for the target data model for seamlessly integrating with developed AI use cases. The tasks include the development of all appropriate ETL/ELT flows and pipelines, the orchestration and CI/CD process, which shall extract the information from the data source, perform the appropriate data transformations (data wrappers, data curation, and data wrangling) and make it available as applicable input to the AI deliverables for each business use case. The pipelines are divided according to the environment and the type of workload they aim to support, so that CI/CD and consequently DataOps and MLOps are feasible.
  - ***WPG.4.1 For the Batch/Stream based approached and per Business Use case***
    - Create the pipelines supporting the low latency ingestion for streaming processing.
    - Create the pipelines supporting the integration/ingestion for batch processing.
    - Create Staging Data pipelines using extracts and delta from the data sources.
    - Create Data Tables for Knowledge exploration applicable for the Data Science process and DEV environment, contributing on the creation of the Machine Learning process and accuracy.
    - Generate Pipeline processes which will load from the Raw Layer and it can apply hard business rules (*which do not change the meaning of the data, just the shape*). A Creation of Metadata Data Tables Loads is also considered. The pipeline process shall capture control flow, data flow, package flow and process metadata and it is used to support error inspection, root cause analysis, performance metrics, timing metrics, performance metrics, volume metrics, error metrics, frequency metrics and dependency chains.

        - Implementation of Data Quality processes via data pipelines shall capture business expectations regarding data quality dealing with data errors, business rules errors and perception errors.
        - Deploy seamlessly pipelines for all environments also performing any connectivity external/internal needed to support the outcomes.
    - ***WPG.4.2 CI/CD:*** The CI/CD process and respective considered sub-processes enabling automation and continuity under the proposed platform hosting the use cases.
        - ***WPG.4.2.1 MLOps:*** Articulate the pipelines created and deploy an automation from DEV to Production environments. Set the KPIs for a successful AI deployment require human intervention, in order to define the best possible set of model parameters.
        - ***WPG.4.2.2 DataOps:*** Enable the practice for each deployed environment, set monitoring and threshold KPIs for initiative-taking action and resolution.
- ***WPG.5 Business Interface Implementation:*** In this activity **Lot1** experts implement the appropriate technical interfaces as exposed by the AI tools layer or via the selected **DSML platform** that makes its data available to its consumers in accordance with the organization's Interface Specification. In addition, in cooperation with the **eu-LISA experts** all appropriate configurations (*platform and network*) are performed (*at first to the DEV and consequently to UAT/QA environments*) to ensure secure exposure of the interfaces to the consuming systems.
- ***WPG.6 Testing.*** Depicts internal and UAT testing leveraging the DEV and UAT/QA environments along with respective EU-LISA's same approach consuming environments. The testing activities include the hosted platform tests and the AI use cases testing in parallel, depicted from the CRISP-DM processes of manual and automation testing:
    - ***WPG.6.1*** Source-to-Target, representation testing to ensure that all transformations are correctly implemented and represented on CDS access layer adequately and correctly.
    - ***WPG.6.2*** Data volumes comparisons to ensure data completeness (all expected rows have reached the virtualization area).
    - ***WPG.6.3*** Performance testing to ensure that data refresh rates are according to the requirements and response times of the AI input information rates are the expected ones.
    - ***WPG.6.4*** Security testing to ensure each AI components access policies are correctly implemented and accessed by similar or other platforms within **eu-LISA**.
    - ***WPG.6.5*** Stress testing to ensure expected performance and availability under extensive load for each implemented use case and hosted-platform service (*in terms of data volumes and in terms of consumer loads*).
    - ***WPG.6.6*** A number of a sub-set or superset of the before mentioned activities in a form of a Test-Object-List mutually agreed with **eu-LISA** as part of a UAT process. This shall explore the appropriate confidence rate that the AI deliverables produce.
- ***WPG.7. Transition to Production:*** After a successful UAT process, project team ensures that deployed changes are declared as SUCCESS on the DEV and consequently on QA/UAT environments and will proceed on selecting a proper cutover window. The transition period along with cutover window to be selected must consider that implemented AI use cases, respective enabling Tools and consuming systems are prepared, selected personnel is available to perform the cutover, affected stakeholders have been notified, and a time window for the process has been scheduled. The process is iterative as it is closely coupled on how possible sub-elements of the use cases and affected systems are independent. After cutover actions, a grace period of early support and system monitoring must be established. For such a purpose it is suggested that a clear contingency plan to exist for each generated aspect of the selected AI use cases (*and/or underlying sub-modules*). For instance, mandatory mitigation actions or recovery options, depending on the identified fault, should exists on how roll back changes may be feasible to be performed in case attempted.

Contract Closing Phase

WPH Closeout: Once the operational scope is finished along with the mutually agreed transition in production, the ***delivery organization must engage and involve necessary project members*** **to assure proper business and system continuity** under the following tasks:

- ***WPH.1 Early Phase Support & Monitoring:*** Delivered work as part of the **AI BUCs**, will follow the applicable standards established by **eu-LISA** to ensure easy maintenance and assure continuity. In more detailed manner, all line of work of newly deliverables or adjustments to existing shall be designed in a way to be easily maintainable and all data will be modifiable without recompilation of the application.
- ***WPH.2* Handover:** A well-planned, properly structured and executed handover of all project objectives and tasks, including analysis, design, implementation, testing, and deployment activities is critical and of major importance. The proposed handover strategy applies the PM2 methodology closing phase guidelines and thus designed in a way to ensure continuity and no disruption or degradation of the services through a controlled transferability. Within the context of this activity, the user acceptance is verified and the project is then capable of being supported from EU-LISA and transfer ownership to respective stakeholders. The handover will be split into three phases, incorporating standard principles at every stage to mitigate potential risks:
  - ***WPH.2.1 Handover activities,*** where all preparation, planning and coordination activities will take place for the delivered use cases, delivered systems and the affected surrounding systems.
  - ***WPH.2.2 Knowledge Transfer,*** where all knowledge and experience will be transferred related to the deliverables, documentation and lessons learned.
  - ***WPH.2.3 Handover Finalization***, where handover is finalized and assessed as per its effectiveness.

8.4 Definition and Quality of the Proposed Project Team

The respective section denotes the proposed team experts able to execute the work described in the indicative tasks having the required seniority, experience and analytical skills as per the potential line of work mandates. The team members that will be explicitly presented in the following lines are proposed to fulfil the anticipated requirements included and identified as part of this study. For every project related to AI and Open-source technology, it is deemed necessary from the authority to put in place an assessment contractor's resources mechanism for a potential pre-screening exercise, a presentation or interviews of the proposed team members as per their preference. The team is proposed to consist of the following roles *(indicative description per role can be found in section §9.4.2)*:

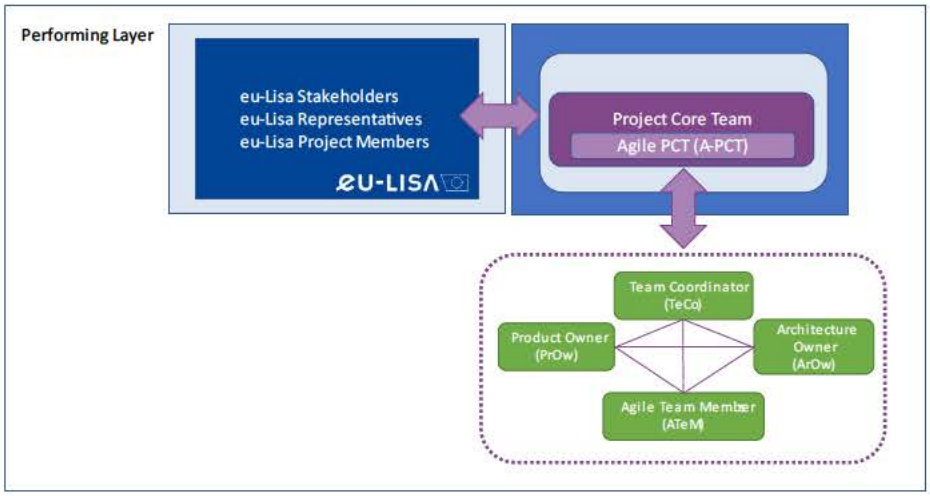


Figure 38. Proposed A-PCT structure

- Project Director/Project Owner (PD)
- Project Manager (PM)
- DPO (Data Protection Officer)
- Agile Project Core Team (A-PCT)

- Product Owner (PO)
- Scrum Master (SM)
- Solution Architect/ Data Expert (SAL)
  - Lead Architect/Architecture Owner
- Agile Team Members (ATeM)
  - Solution Architect/ Data Expert (SA)
  - The AI Consultant (CO)
  - Interface Designer (ID)
  - IS Tester Expert (TE)
  - DevOps Engineer (DV)
  - Business Analyst (BA)
  - System Engineer (SE)
  - Data Engineer (DE)
  - Data Scientist (DS)

Based on the mentioned methodologies and particularly with selected PM2 methodology, a robust framework must be developed, including *initiative-taking prevention, real-time monitoring, corrective actions, and continuous efficiency improvement*. A proposed governance model *(following also an Agile PM2 approach)*. **Agile PM2** may be adopted introducing an Agile Project Core Team *(A-PCT)* under the following team member roles: **PrOw**: *The 'single voice of the stakeholders' within the Agile Project Core Team acting as the Product owner*. Should work, if possible, within the physical/virtual space of the team. **TeCo**: *Facilitator and team coach whose main purpose is to allow the team to focus on achieving specific objectives and being successful, can be covered by a Scrum Master as he synchronizes also with the PM*. **ArOw**: *Responsible for the architecture decisions for the Agile Project Core Team (A-PCT), which is typically the Lead Architect.* Leads the initial architecture envisioning effort at the beginning of the project and ensures alignment with the guidelines and recommendations of the Architecture Office (AO). **ATeM**: A team that implements the solution.

8.4.1 Project Governance Framework

To assure the effective and efficient execution of each anticipated task and the activities issued under its context in conformity with the requirements and planning, a governance methodology for similar implementations is being proposed (PM2) as can be seen on the figure below.

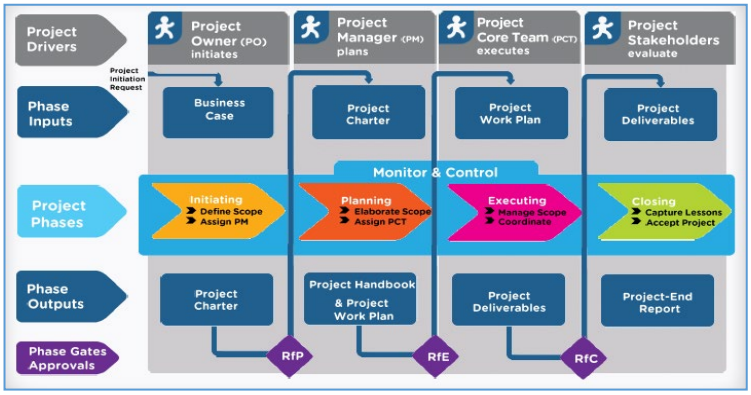


Figure 39. Project Governance framework

Under this concept, monitor and control of the contracted tasks and activities, is applied on all project phases. The Monitor & Control shall run in parallel for all associated project phases and contracted tasks. The Project Manager will be responsible for monitoring the project's execution and assessing the evolution of work, assuring the feasibility of the delivery, in accordance with the schedule and workload estimates defined. To enable the thorough assessment of the Project's evolution, our proposal is to the potential project to entail the following:

- Adoption of necessary phase gates *(i.e. RfP, RfE, RfC)* as check points to contribute to overall management quality for each handled task under each selected use case as part of the scope.
- Establishment of specific milestones in accordance with the envisaged execution of the project's tasks. This will facilitate a regular assessment of the project's evolution and identify elements which deviate from the initial estimations (schedule / workload) or even deliverables of this study. Milestones enable to identify issues as early as possible in the process where corrective actions can be easier implemented, ensuring that the agreed deadlines are not impacted.
- The need to define prior and specific metrics (KPIs) enabling the continuous assessment of project evolution on various elements, namely: quality (deliverables and underlying code); punctuality; efficient use of resources; knowledge acquisition, recording and dissemination. This is deemed necessary as to be able to identify deviation from the initial planning and scope definitions of the use cases and revert accordingly.
- Adoption of a continuous feedback acquisition assessment loop leading to the prompt identification and understanding of imminent issues. This involves specific **eu-LISA's** stakeholders. In this way, the potential contractor shall be able to take all necessary actions to eliminate them as early as possible, before affecting the project's execution.
- Assurance of effective and regular communication and reporting (internal and external) between all involved stakeholders.
- Facilitation of knowledge acquisition, recording, maintenance, and dissemination through the execution of the project, and during the handover and takeover periods.
- Identification, monitoring and handling of all relevant risks that can impact execution in any way.
- Produce and deliver artefacts based on the respective phases and drivers of each contracted task.

For the establishment of the AI BUCs as part of a single project, for **eu-LISA**, we foresee the governance framework to contain the following key groups of users:
The A-PCT produces the project deliverables and plays a key role in the planning and the successful completion of the project. In our approach this team is enhanced under Agile principles and is described in main sections as A-PCT team.
The PO represents the business side of the project, chairs the PSC, and is accountable for the overall project success. The PO typically holds a management role in the functional organization.
The PM manages the project and assumes responsibility for the project and project management work and deliverables.

- The Project stakeholders, steers and interacts with the project lifecycle based on the use cases and underlying tasks or activities. Additionally, they are the ones for signing off respective key artefacts and deliverables of the foreseen project.

Manages the project and assumes responsibility for the project and project management work and deliverables

8.4.2 Foreseen Profiles

The recommended profiles based on our approach and as needed to support the establishment of the identified business use cases, for the respective responsibilities and roles are described as follows:

The ***Project Director/Project Owner (PD)*** is responsible for:
Ensuring availability of all resources necessary for the execution of the contract;
Follow-up the proper execution of the Project Contract;
Will function as the related Business Manager covering the business aspects of the project as per Project Management methodology principles;
Serves as an escalation point for any issue that may arise in the lifetime of the contract.

The role of ***Project Manager (PM)*** will be responsible for the whole project, support and maintenance planning KPIs. Their responsibilities include:
Monitor and control progress of all envisaged tasks and activities ensuring compliance to the defined schedule, the adopted Project Management and Software Development methodologies and the agreed Service Levels (in case of Operational Support);
Ensure the effective management of Problems and Change Management;
Report to stakeholders about the status of the project;
Facilitate the provision of necessary support to address performance issues of the Project Team members;
Ensure the continual improvement in all aspects of the project.
Single Point of Contact (SPoC) with the Customer, ensuring the efficient communication through formal (the agreed meetings) and informal means;

Report progress, in accordance with the defined reporting requirements;
Responsible of Change Management and Risk Management.

The role of ***Project stakeholders*** will be responsible for the whole project, support and maintenance planning KPIs.

It is suggested to include and assign certain EU-Lisa's key stakeholders that will support the decision making and are affected by the introduction of these use cases.

The role of **A-PCT** will be responsible for the establishment of the whole project. Their responsibilities include:
The AI Consultant will provide assistance to the Project Manager and the team members considering vertical aspects of the project, i.e. related to business and technology. Certain role is considered as a Subject Matter Expert and shall assists in requirement analysis, design, project takeover and handover. Their responsibilities include:
Participate in various project phases with emphasis in project take-over and handover;
Business analysis based on the subject matter expertise, i.e. AI, IT, Big Data Analytics;
Thematic areas of CRRS, ETIAS etc;
Provision of consultancy and support to the project team members, during established tasks;
Serve as a subject matter expert for the education aspects or IT aspects of the project;
Ability to analyse to the stakeholders the requirements and design and implement appropriate solutions.

The Solution Architect/ Data Expert (SA) will be responsible in reviewing and defining relevant architecture and articulate thorough analysis activities. Respective resources shall provide assistance to the Project Manager for the production of main and sub-tasks and shall be responsible for the analysis and design of changes and the update of Analysis & Design artefacts. Their responsibilities include:

- Drive and define a data architecture vision and strategy to maximize the optimization of data assets;
- Review and understand the architecture design of the system (and its submodules);
- Analyse requirements, translate them into architectural designs and specifications for development;
- Analyse and model the underlying data, map data sources, document interfaces and data movement;
- Coordinate / provide expertise during the implementation of the relevant technical architecture.
- Identifying and documenting business processes, issues, value levers, metrics, and capabilities;
- Completing analysis and documentation of as-is and to-be capabilities;
- Update the relevant documentation, create instructions and reports (e.g. test reports), and prepare mock-ups for structure validation;
- Analyse and model the underlying data;
- Coordinate / provide expertise during the implementation of the relevant technical architecture;
- Analyse, understand and model business processes;
- Analyse requirements and translate them into functional designs and specifications for development;
- Update the relevant documentation, create instructions and reports (e.g. test reports), and prepare mock-ups for structure validation;
- Participate in system and acceptance testing activities;
- Produce and maintain documentation (user and technical manuals, etc.) as well as online help material during the implementation process.
- Analyse new development in the required tasks, to understand how they will impact and can be incorporated in existing systems;
- Update the relevant documentation, create instructions and reports (e.g. test reports), and prepare mock-ups for structure validation;
- Provide knowledge sharing and handover activities.
- Maintain documentation (user and technical manuals, etc.) as well as online help material, in accordance with the performed updates.

The ***Interface Designer (ID)*** will be responsible for the analysis of the business, the functional and non-functional requirements and their transformation into graphical interfaces and solution designs.

- Provide analysis, design, testing, implementation and ongoing support for managing interfaces;
- Responsible for driving the foreign system interface design process;
- Coordinate analysis and design sessions; document specifications into an Interface Control Document (ICD);
- Supports the design and management of systems over the entire lifecycle from inception to phase-out.
- Prototyping Graphical User Interfaces (wireframes, mock-ups, etc.);
- Design and implement data models to support reporting, dashboarding, and integration needs;
- Production of Graphical User Interface; Installs, configures, tests, and maintains operating systems, applications software, and system management tools.

The ***IS Tester Expert (TE)*** will perform and participate under the following tasks:

- Perform in system and acceptance testing activities;
    - Organize and execute regression testing;
    - Support the acceptance testing activities;

- Execute unit and integrations tests;
  - Participate in incident/problem handling and change implementation activities;
  - Produce and maintain documentation (technical specifications, and user guides, etc.) as well as online help material during the implementation process;
  - Organization and provision of all training activities;
  - Provision of technical assistance and support.
- Production of test plans and test cases;
- Organize and execute regression testing;
- Support the acceptance testing activities;
- Report testing results;
- Provision of support during the handling of the issues identified during testing.

The ***DevOps Engineer (DoE)***, who have seniority and particular expertise in the CI/CD technologies requested to support all use cases shall be responsible for the following:

- Automate the Pipeline and release processes;
- Enable and develop along with other project resources the MLoPS and DataoPS activities;
- Support the deployment in all environments of eu-LISA;
- Participate in system and acceptance testing activities;

The ***Business Analyst (BA)*** will be responsible for acting as a liaison between EU-Lisa's business units and data warehouse technical development team to ensure business requirements are understood and translated into functional and technical requirements. The Business Analyst will collaborate with stakeholders from EU-Lisa's business units to gather and record user requirements for the AI tools (use cases) enablement. They will closely work with the project team to ensure user requirements are understood before and during the development phase. The Business Analyst, along with the Project Manager will be responsible to ensure processes are defined, documented and translated into technical requirements for the development teams to build and deliver the requirements. Main responsibilities include:

Work with projects' stakeholders to ensure that requirements are clearly defined, delivered and approved for deployment.

Translate business requirements into detailed system requirements and technical design specifications.

Conduct requirements analysis for change requests including requirements gathering

Participate in sprint planning

The ***System Engineer (SE),*** shall analyse the requirements of the system, takes part in architectural design, perform infrastructure implementation, as well as the design documentation and execution of acceptance testing. The responsibilities of this project role include:

- Analysis and documentation of system requirements, design on parts of architecture and initiation of subsystem requirements.
- Operating system, network, storage, virtualization and containerization configuration.
- System software (operating systems, databases), installation, configuration, and adaptation.
- Verification that subsystems are implemented in line with system requirements and architecture.
- Planning, documentation, and execution of system testing.
- Optimization of system performance.
- Identify improvement opportunities as well as adopting new methods and best practices for eu-LISA.

The ***Data Engineer (DE)*** will be responsible for reviewing, defining and implementing relevant pipelines and articulate thorough analysis activities. Respective resources shall assist the Project Manager for the production of main and sub-tasks, following responsibilities are considered:

Analyse and model the underlying data, map data sources, document interfaces and data movement.

Coordinate / provide expertise during the implementation of the relevant technical architecture.

Analyse, understand, and model pipeline processes.

Implement ETL/ELT.

Implement Data Integration and Data Ingestion pipelines.

Participate in system and acceptance testing activities.

Analyse new development in the required tasks, to understand how they will impact and can be incorporated in existing systems.

Provide knowledge sharing and handover activities.

The ***Data Scientist (DS),*** are primarily responsible for using various programming algorithms so that they can build, test and deploy the models necessary to support the defined business use cases, following responsibilities are considered:

Apply data science techniques, such as machine learning, statistical modelling, and artificial intelligence;

Creating and managing the AI/ML development process and overall infrastructure of each use case;

Transform the machine learning models into APIs that other applications can interact with;

Perform and articulate outcomes of the Exploratory analysis for each use case;
Identifying relevant data sources for the business use cases and apply quality mechanisms to support adequate model processing;
Implement Models and compose deliverables applicable for the provisioning to the AI tools platform (i.e., DSML)

8.4.3 Indicative Time Plan

This section provides an overview of the indicative plan that the project may be based upon. The project phases are closely coupled with the proposed PM2 methodology, thus and as per section §9.3, they consist of: the Contract Initiation phase, the contract Planning phase, the contract execution phase and the Contract Closing phase. The plan is meant to provide a high-level estimate on how the artefacts of this study may be established as part of a number of project tasks and deliverables. The proposed duration is 194 days with an approximate of 24 Sprints executed sequentially or upon certain activities in parallel. The following figures describe in detailed the proposed project plan also considering the delegation of activities per phase and tasks.

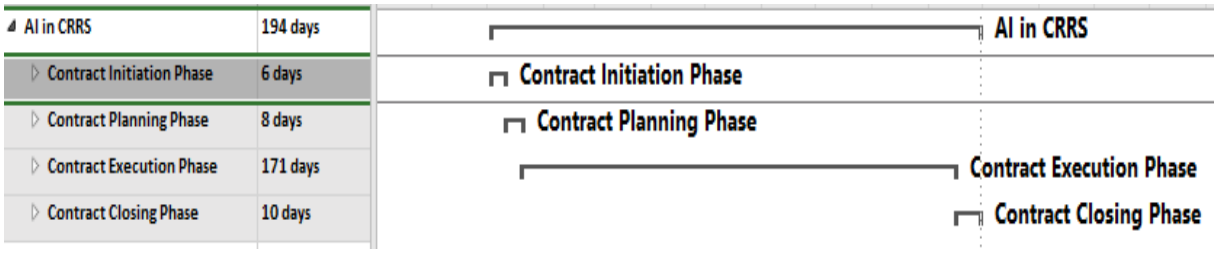


Figure 40. High-level overview of the Proposed Plan

Here follows the indicative schedule for the Contract Initiating phase:

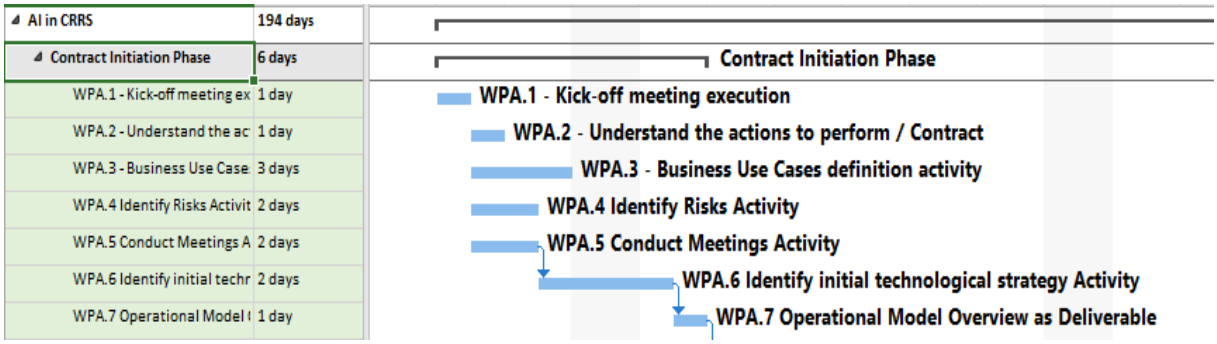


Figure 41. Overview of the Contract Initiation Phase

The indicative schedule for the Contract Planning phase is displayed in the following figure:

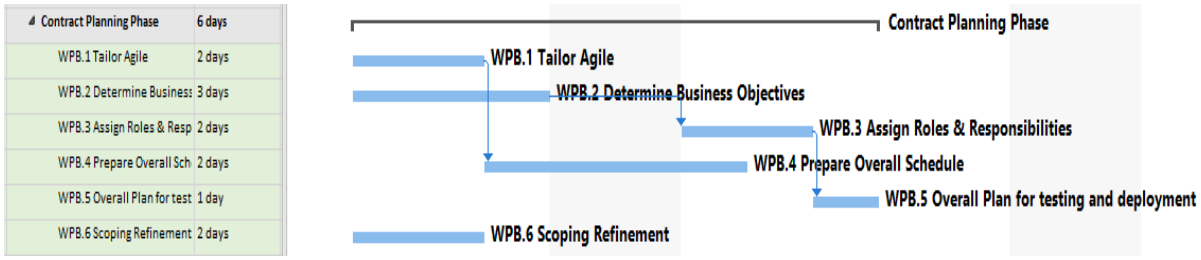


Figure 42. Overview of the Contract Planning Phase

The indicative schedule for the Contract Execution phase is displayed on **Figure 53**. The displayed schedule is indicative and includes seven MVPs (as the conducted study) and respective project establishment. The work-packages described earlier under this section are being separated into a series of 2-week sprints executed either in parallel or sequentially according to the line of work anticipated to be delivered.

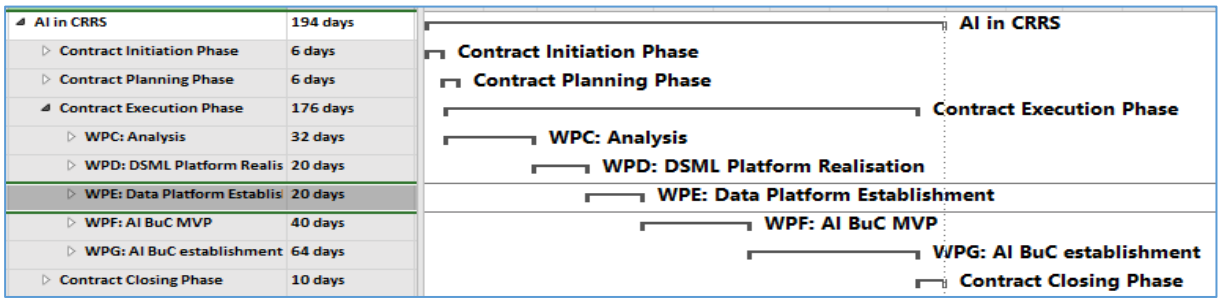


Figure 43. High-Level overview of the Contract Execution Phase

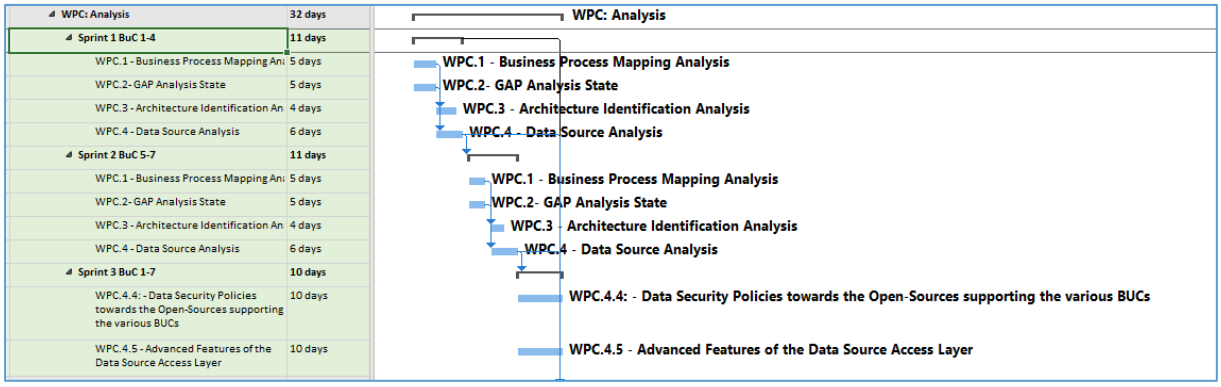


Figure 44. Contract Execution Phase: WP.C Analysis

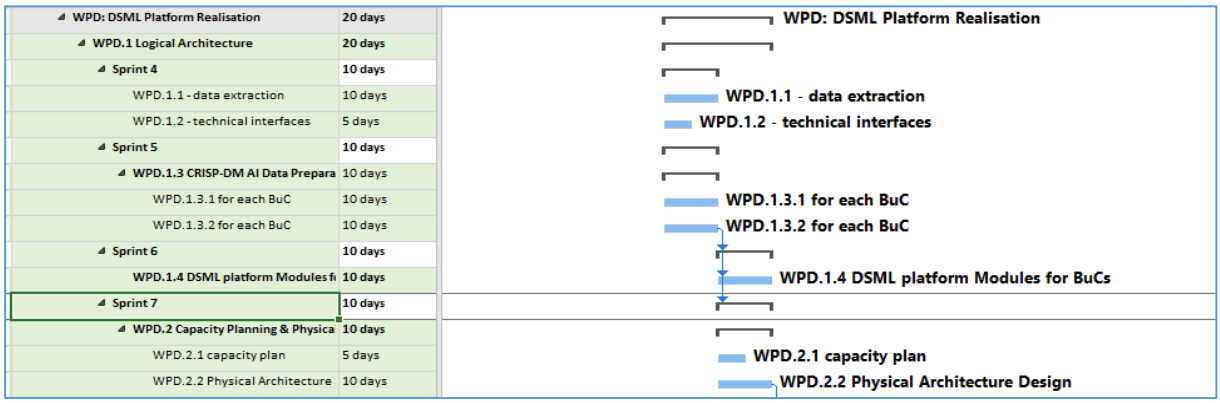


Figure 45. Contract Execution Phase: WP.D DSML Platform Realisation

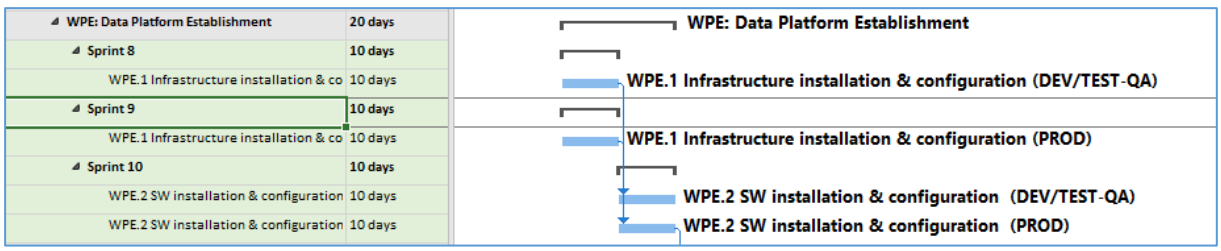


Figure 46. Contract Execution Phase: WPE Data Platform Establishment

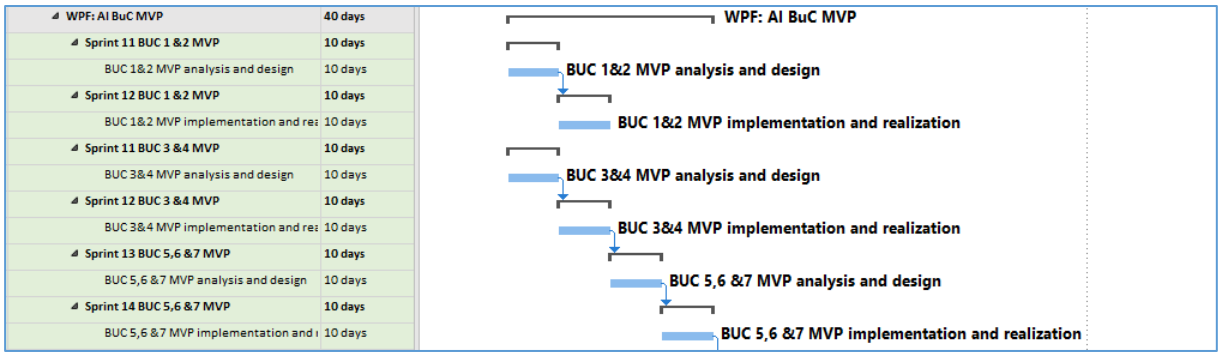


Figure 47. Contract Execution Phase: WPF AI BUC MVPs

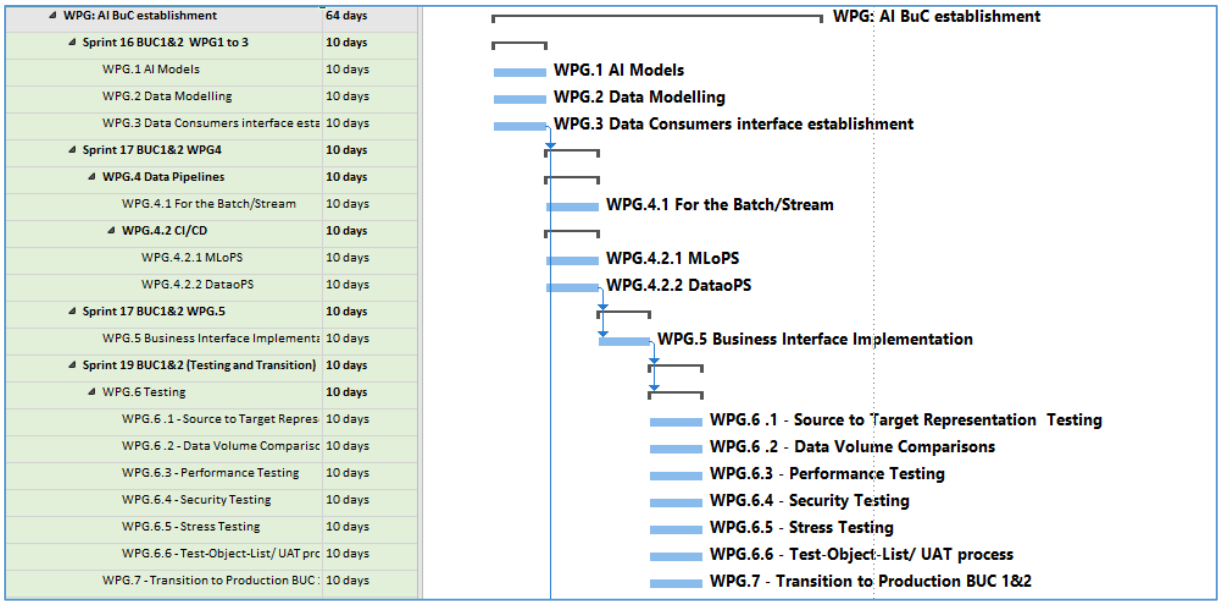


Figure 48. Contract Execution Phase: WPG AI BUC establishment for BUC 1&2

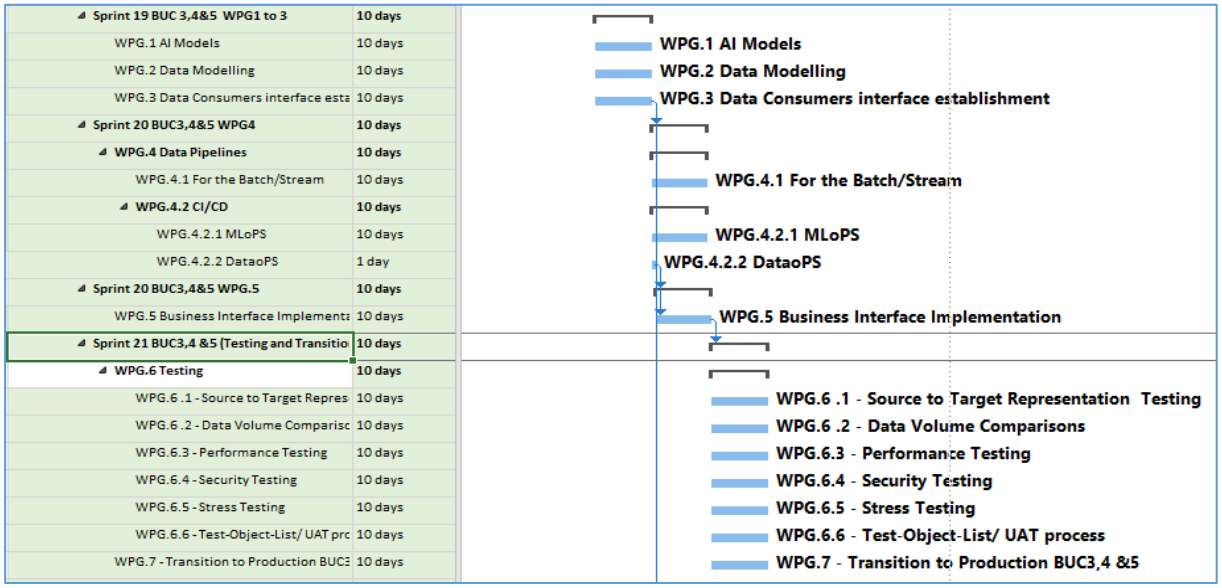


Figure 49. Contract Execution Phase: WPG AI BUC establishment for BUC 3,4&5

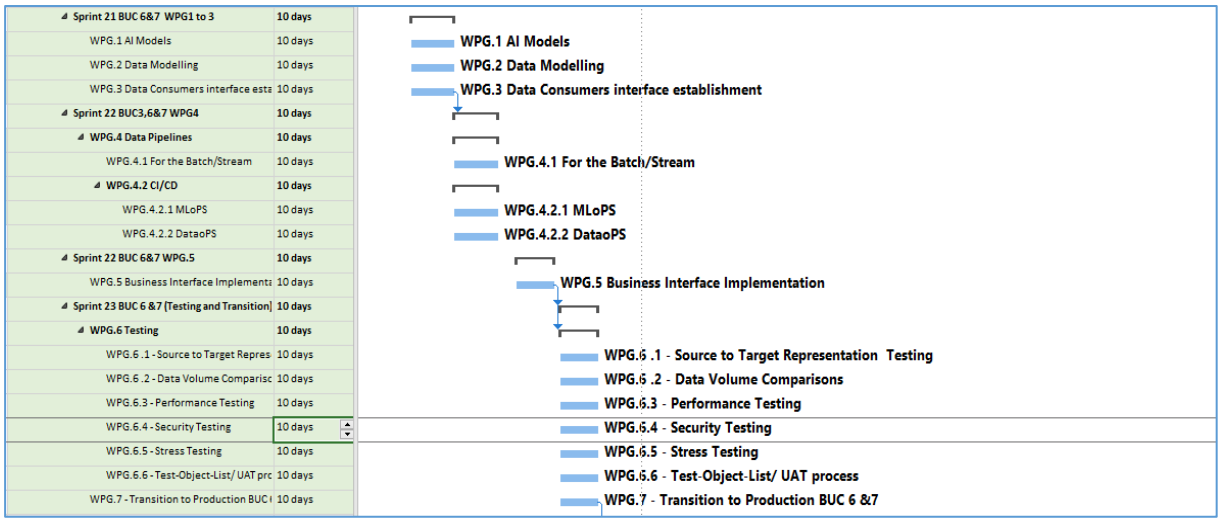


Figure 50. Contract Execution Phase: WPG AI BUC establishment for BUC 6&7

The indicative schedule for the Contract Closing phase is displayed in the following figure:

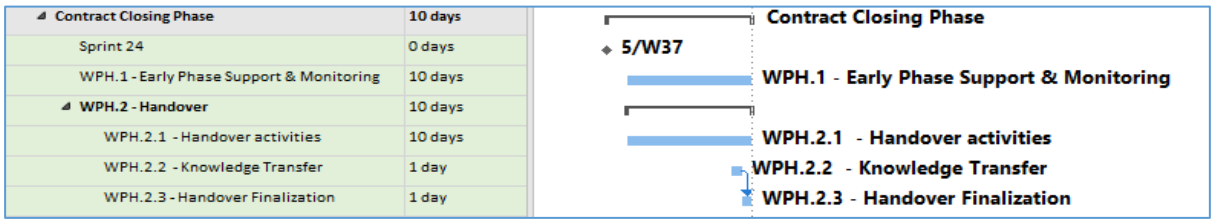


Figure 51. Contract Closing Phase

8.4.4 Indicative Effort Scheme and allocations

This section encompasses a high-level estimate of the anticipated Man-days per profile according to the Three-point estimate technique *(PERT)*. The PERT technique may be a starting point that can help determine the expected schedule, and implicit costs for the anticipated project. The three-point estimate uses three values, to establish an overall estimate for the cost or time of a project and its components. For the shake of clarity, three distinct estimates will be generated according to project phases and work-packages measured in total Man-days. The estimate will then be delegated per anticipated role per phase or per sub activity. The three estimates to be provided are described as follows:

- Optimistic estimate (O): Corresponds to the best-case scenario, considering the least amount of time or effort a project or activity would need.
- Most likely estimate (M): Considers the average or usual amount of time or effort the project team would need to establish the project.
- Least likely estimate (L): in the worst-case scenario, if everything somehow goes wrong, how long or how much effort a project or activity might need.

The following table describes the estimate as was extracted from the indicative effort estimation per role:

	WBS								
	Approach 1 in Man-days	Approach 2 in Man-days	Approach 3 in Man-days						
Component #	**Expected**	**Optimistic**	**Pessimistic**	**PERT Value**					
WPA Initiating	19,20	15,36	34,56	21		PERT Estimate = (4*Expected Estimate + Optimistic Estimate + Pessimistic Estimate)/6			
WPB Planning	22,20	17,76	39,96	24					
WPC Analysis	195,52	156,416	351,936	215					
WPD DSML Platform Realization	104,00	83,2	187,2	114		**Optimistic**	**1598,61 Man-days**		
WPE Data Platform Establishment	78,00	26	140,4	80		**Expected**	**2043,76 Man-days**		
WPF AI BuC MVPs	488,40	390,72	879,12	537		**Pessimistic**	**3678,77 Man-days**		
WPG AI BuC establishment	1069,44	855,552	1924,992	1176					
WPH Closeout	67,00	53,6	120,6	74					
Total	2043,76	1598,608	3678,768	**2242**					

Table 34. PERT approach with O-M-L estimate in Man Days per described phase of the anticipated project.

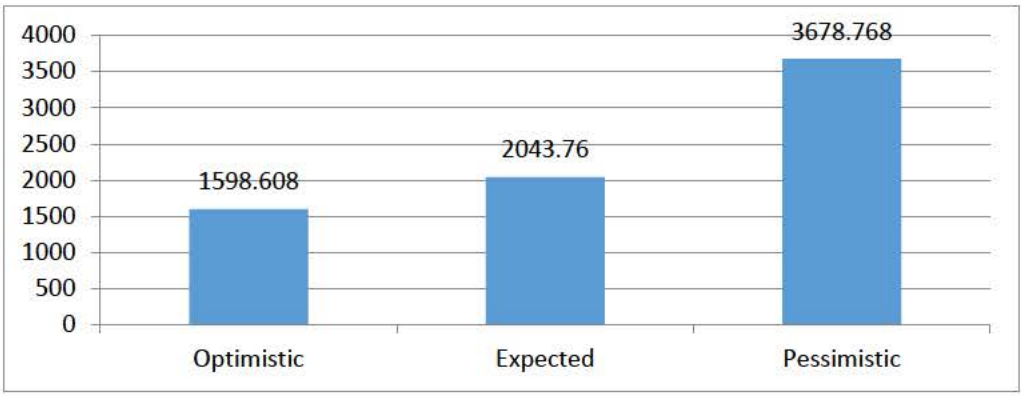


Figure 52. Produced results for the three-point estimate

The anticipated average of FTEs is being calculated as a number of almost 10 resources. An indicative allocation per profile that was used to derive the results of

	WBS								
	Approach 1 in Man-days	Approach 2 in Man-days	Approach 3 in Man-days						
Component #	**Expected**	**Optimistic**	**Pessimistic**	**PERT Value**					
WPA Initiating	19,20	15,36	34,56	21		PERT Estimate = (4*Expected Estimate + Optimistic Estimate + Pessimistic Estimate)/6			
WPB Planning	22,20	17,76	39,96	24					
WPC Analysis	195,52	156,416	351,936	215					
WPD DSML Platform Realization	104,00	83,2	187,2	114		**Optimistic**	**1598,61 Man-days**		
WPE Data Platform Establishment	78,00	26	140,4	80		**Expected**	**2043,76 Man-days**		
WPF AI BuC MVPs	488,40	390,72	879,12	537		**Pessimistic**	**3678,77 Man-days**		
WPG AI BuC establishment	1069,44	855,552	1924,992	1176					
WPH Closeout	67,00	53,6	120,6	74					
Total	2043,76	1598,608	3678,768	**2242**					

Table 37, follows:

		FTEs						
AI in CRRS	**Duration (days)**	**PD**	**PM**	**PO**	**SM**	**SAL**	**SA**	**CO**
Contract Initiation Phase	6	0,10	0,60	0,10	0,10	0,90	0,20	0,20
Contract Planning Phase	6	0,10	0,60	0,10	0,10	1,00	0,20	0,60
Contract Execution Phase	176							
WPC: Analysis	32	0,01	0,10	0,20	0,20	1,00	2,00	1,00
WPD: DSML Platform Realisation	20	-	0,10	0,20	0,20	0,70	2,00	0,50
WPE: Data Platform Establishment	20	-	0,10	0,20	0,20	0,50	1,00	-
WPF: AI BuC MVP	40	0,01	0,30	0,50	0,50	0,70	2,00	1,00
WPG: AI BuC establishment	64	0,01	0,20	0,50	0,50	0,50	2,00	0,50
Contract Closing Phase	10	0,20	0,20	0,10	0,10	0,10	0,80	1,00
AI in CRRS	**Duration (days)**	PD	PM	PO	SM	SAL	SA	CO
Contract Initiation Phase	6	0,60	3,60	0,60	0,60	5,40	1,20	1,20
Contract Planning Phase	6	0,60	3,60	0,60	0,60	6,00	1,20	3,60
Contract Execution Phase	176							
WPC: Analysis	32	0,32	3,20	6,40	6,40	32,00	64,00	32,00
WPD: DSML Platform Realisation	20	0,00	2,00	4,00	4,00	14,00	40,00	10,00
WPE: Data Platform Establishment	20	0,00	2,00	4,00	4,00	10,00	20,00	0,00
WPF: AI BuC MVP	40	0,40	12,00	20,00	20,00	28,00	80,00	40,00
WPG: AI BuC establishment	64	0,64	12,80	32,00	32,00	32,00	128,00	32,00
Contract Closing Phase	10	2,00	2,00	1,00	1,00	1,00	8,00	10,00

Table 35. FTE allocation per activity of lead A-PCT team members

		FTEs						
AI in CRRS	**Duration (days)**	ID	TE	DV	BA	SE	DE	DS
Contract Initiation Phase	6	-	-	-	1,00	-	-	-
Contract Planning Phase	6	-	-	-	1,00	-	-	-
Contract Execution Phase	176							
WPC: Analysis	32	0,10	-	-	1,50	-	-	-
WPD: DSML Platform Realisation	20	-	-	-	1,50	-	-	-
WPE: Data Platform Establishment	20	-	0,20	0,20	-	1,50	-	-
WPF: AI BuC MVP	40	0,50	-	-	1,50	0,20	2,50	2,50
WPG: AI BuC establishment	64	2,00	1,00	2,00	1,00	1,50	2,50	2,50
Contract Closing Phase	10	0,10	-	-	-	0,10	1,00	3,00
AI in CRRS	**Duration (days)**	ID	TE	DV	BA	SE	DE	DS
Contract Initiation Phase	6	0,00	0,00	0,00	6,00	0,00	0,00	0,00
Contract Planning Phase	6	0,00	0,00	0,00	6,00	0,00	0,00	0,00
Contract Execution Phase	176							
WPC: Analysis	32	3,20	0,00	0,00	48,00	0,00	0,00	0,00
WPD: DSML Platform Realisation	20	0,00	0,00	0,00	30,00	0,00	0,00	0,00
WPE: Data Platform Establishment	20	0,00	4,00	4,00	0,00	30,00	0,00	0,00
WPF: AI BuC MVP	40	20,00	0,00	0,00	60,00	8,00	100,00	100,00
WPG: AI BuC establishment	64	128,00	64,00	128,00	64,00	96,00	160,00	160,00
Contract Closing Phase	10	1,00	0,00	0,00	0,00	1,00	10,00	30,00

Table 36. FTE allocation per activity of ATeM team members

8.4.4.1 Foreseen indicative Risks

Risks hindering the provision of high-quality services can emerge during the provision of the envisaged use cases. In this context, a Risk Management Approach based is proposed to be established and used. Using this approach, a detailed Risk Management plan shall be created as part of the Planning Phase. The plan is usually performed at the commencement of the contract, which must be revalidated at the commencement of each AI business use case, as it is a cyclic process (Monitor & Control process, Project Logs, Risk Log). Risk management is a cyclic process, thus the risks are monitored and mitigated after proper actions within the duration of the project. A short list of indicative risks and their proposed mitigation actions, identified as per past experiences and prior workshops information extraction, follows:

No.	Area	Risk	Mitigation Actions	Likelihood (1-5)	Impact (1-5)
1	Technical and Organizational	Delays due to dependencies with	Avoid conflicts between projects; Manage dependencies; Capacity planning (teams,	3	3

No.	Area	Risk	Mitigation Actions	Likelihood (1-5)	Impact (1-5)
		other projects	resources, technological environment etc); Daily meetings between teams; Overview by PM; Follow product development approach.		
2	Technical and Organizational	Delays due to dependencies between A-PCT teams and key activity stakeholders	Manage dependencies; Capacity planning; Daily scrums between teams; Overview by PM and PD; Follow product development approach	3	3
3	Technical and Organizational	Lengthy, tedious, and complex process of finalizing User and functional requirements.	Perform a-priori analysis and produce necessary risk evaluation artefacts. Define the scope and functional requirements via workshops with respective stakeholders/SMEs to receive proper input.	3	3
4	Technical and Organizational	Project Schedule /or sub- Tasks and activities schedules are not clear	Hold frequent meetings with the A-PCT team, to understand plan and respective tasks, receive input from stakeholders, reduce the likelihood to not complete in-time activities.	2	3
5	Organizational	Lack of commitment of eu-LISA' or contractor's teams due to overflow in Day-by-Day activities	Strong Change Management aimed to improve commitment of Business (Top Management included);Team building activity involving; business, IT and external suppliers; Definition of SPoC for each task	2	4
6	Organizational	Stakeholders/Business users not always easy to identify or not involved or committed actively	Use of proper methodology to identify; Stakeholders through specific selection criteria; Create stakeholders' matrix and expected involvement from ordering process; Early stakeholder involvement; Collection of feedback from stakeholders during project activities	2	5
7	Organizational	A large number of different stakeholders in eu-LISA impacts the project, the more likely it is that various conflicts will arise.	Effective management and early involvement of project stakeholders to necessary phases and milestones becomes a very important task for project success and mitigate accordingly risks of failures or delays.	4	5
8	Technical	Extending 3rd Party code. Low quality/undocumented on previous deployment of PoCs.	Perform necessary inception workshops to take-over knowledge from respective stakeholders. Use S.M.Es and Senior personnel to document necessary information in a form of artefacts.	2	2
9	Technical	Low quality in solution deployment	Development methods evolving to Agile; Use of Task Automation Tools; Use of DevOps; Use of Data Analytics established principles.	2	4

No.	Area	Risk	Mitigation Actions	Likelihood (1-5)	Impact (1-5)
10	Technical	Takeover on existing Solutions/Platforms for support resolution	A number of take-over familiarization activities and knowledge exchange/workshop with respective system's stakeholders should be performed to fasten the process of take-over and assure business continuity	2	4
11	Technical	Increase of resources demand, problem with key resources – unforeseen unavailability	We assure a knowledge transfer is performed though out all our team and project member assigned under this contract. We aim to ensure there is knowledge redundancy among the members that even in unforeseen unavailability of a specific team member, the remaining staff can undertake his/her tasks successfully.	2	3
13	Financial	Non-realistic economic effort estimation	Proper management of Agile projects; Use agreed methodology	2	4
14	Business Change	New requirements or changes in scope due to dependencies among projects	Agile approach; Use of common infrastructure and configuration tools; Risk & Quality function ensures that the TLs of each activity keeps track of the requirements	4	3
15	Business Change	New requirements or changes in scope of baseline	Use of specific PM and agile methodologies that prioritize projects and requirements according to their fit with business needs	4	3
16	External	Changes in regulatory environment (e.g. of IP/Trademarks)	Monitor the work ;Improve tactical level governance; Bring knowledge from other contracts	3	3
17	External	New Technological solutions that could re-shape the market in Open source	Continuous monitoring of benchmark and technology market evolution	2	3

Table 37. Indicative Risks and Mitigation actions

9 Appendix III – Business and Stakeholder Requirements

9.1 RP.01 Identification of risk for a specific group of travellers.

9.1.1 Business Requirements (RP.01)

9.1.1.1 REQ-01.01 (BR): The CRRS shall produce reports and statistics from VIS data.

RP Specific ID	REQ-01.01 (BR)
Requirement ID	AIC-REQ-00009
Name	The CRRS shall produce reports and statistics from VIS data.
Description	The CRRS shall produce reports and statistics from the following VIS data: 1. status information; 2. the competent visa authority, including its location; 3. current nationality of the applicant; 4. Member State of first entry; 5. date and place of the application or the decision concerning the visa; 6. the type of visa issued; 7. the type of the travel document; 8. the grounds indicated for any decision concerning the visa or visa application; 9. the competent visa authority, including its location, which refused the visa application and the date of the refusal; 10. the cases in which the same applicant applied for a visa from more than one visa authority, indicating these visa authorities, their location and the dates of refusals; 11. main purpose(s) of the journey; 12. the cases in which the fingerprints of the applicant could factually not be provided; 13. the cases in which the fingerprints of the applicant were not required to be provided for legal reasons; 14. the cases in which a person who could factually not provide the fingerprints was refused a visa.
Status	Proposed
Source	01: 02008R0767-20210803 VIS REG 02: 02019R0817-20210803 Interoperability Visa Borders REG v1

RP Specific ID	REQ-01.01 (BR)
Rationale	01: Art. 17(1) 02: Art. 39(1)
Keywords	VIS; data; reports; ad-hoc; statistics; information
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.1.2 REQ-01.02 (BR): The CRRS shall provide reports and statistics on EES data.

RP Specific ID	REQ-01.02 (BR)
Requirement ID	AIC-REQ-00003
Name	The CRRS shall provide reports and statistics on EES data.
Description	The CRRS shall provide reports and statistics on the following EES data: a) status information; b) nationality, sex and year of birth of the third country national; c) date and border crossing point of the entry to a Member State and date and border crossing point of the exit from a Member State; d) the type of the travel document and the three-letter code of the issuing country; e) the number of persons identified as overstayers, the nationalities of persons identified as overstayers and the border crossing point of entry; f) the data entered in respect of any stay revoked or any stay whose

RP Specific ID	REQ-01.02 (BR)
	validity is extended; g) the three-letter code of the Member State that issued the visa, if applicable; h) the number of persons exempt from the requirement to give fingerprints (children under the age of 12 or persons for whom is physically impossible to give fingerprints). i) the number of third-country nationals refused entry, the nationalities of third-country nationals refused entry, the type of border (land, air or sea) of the border crossing point at which entry was refused and the reasons for which entry has been refused. The CRRS shall allow access to the duly authorised staff of the European Border and Coast Guard Agency for the purpose of conducting risk analysis, vulnerability assessments and monitoring migratory flows.
Status	Proposed
Source	01: 02017R2226-20210803 EES REG
Rationale	01: Art. 63(1), 63(2)
Keywords	EES; FRONTEX; data; risk analysis; vulnerability assessment
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.1.3 REQ-01.03 (BR): The CRRS shall produce statistics on ETIAS data.

RP Specific ID	REQ-01.03 (BR)
Requirement ID	AIC-REQ-00005
Name	The CRRS shall produce statistics on ETIAS data.
Description	The CRRS shall produce statistical reports from the following ETIAS data: a) application status information; b) nationalities, sex and year of birth of the applicant; c) the country of residence; d) education (primary, secondary, higher or none); e) current occupation (job group); f) the type of the travel document and three-letter code of the issuing country; g) the type of travel authorisation and, for a travel authorisation with limited territorial validity on humanitarian grounds, for reasons of national interest or because of international obligations, a reference to the Member state(s) issuing the travel authorisation with limited territorial validity; h) the validity period of the travel authorisation; and i) the grounds for refusing, revoking or annulling a travel authorisation.
Status	Proposed
Source	01: 02018R1240-20210803 ETIAS REG
Rationale	01: Art. 84(1), 84(2)
Keywords	ETIAS; statistics; data; travel authorisation; FRONTEX
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] Lot1 Contractor)
Final approver	eu-LISA Programme Manager

RP Specific ID	REQ-01.03 (BR)
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.1.4 REQ-01.04 (BR): The ETIAS IS shall store and provide access to data for the purpose of reporting and statistics without allowing for individual identification.

RP Specific ID	REQ-01.04 (BR)
Requirement ID	AIC-REQ-00008
Name	The ETIAS IS shall store and provide access to data for the purpose of reporting and statistics without allowing for individual identification.
Description	The duly authorised staff of the competent authorities of MSs, the Commission, eu-LISA and the ETIAS CU shall have access to consult the following data, solely for the purposes of reporting and statistics, without allowing for individual identification and in accordance with the safeguards related to non-discrimination rights: (a) application status information; (b) nationalities, sex and year of birth of the applicant; (c) the country of residence; (d) education (primary, secondary, higher or none); (e) current occupation (job group); (f) the type of the travel document and three letter code of the issuing country; (g) the type of TA and, for a TA with TV on humanitarian grounds, a reference to the MSs issuing the TA (with LTV); (h) the validity period of the TA and (i) the grounds for refusing, revoking or annulling a TA. For those purposes, eu-LISA shall store the data referred to in that paragraph in the CRRS. The cross-system statistical data and analytical reporting shall allow the authorities listed above to obtain customisable reports and statistics, to support the implementation of the ETIAS screening rules, to improve the assessment of the security, illegal immigration and high epidemic risks, to enhance the efficiency of border checks and to help the ETIAS CU and the ETIAS NUs process TA applications. The procedures put in place by eu-LISA to monitor the development and the functioning of the ETIAS IS shall include the

RP Specific ID	REQ-01.04 (BR)
	possibility to produce regular statistics for ensuring that monitoring.
Status	Proposed
Source	01: 02018R1240-20210803 ETIAS REG 02: 02019R0817-20210803 Interoperability Visa Borders REG v1
Rationale	01: Art. 84(1), 84(2), 84(3), 84(4) 02: Art. 61(16)
Keywords	anonymised; data; identity; TA; decision; CRRS; daily statistics
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.1.5 REQ-01.05 (BR): The CRRS shall provide access to MID, CIR and ESP data stored in CRRS to the EBCGA for the purpose of carrying out risk analyses and vulnerability assessments.

RP Specific ID	REQ-01.05 (BR)
Requirement ID	AIC-REQ-00004
Name	The CRRS shall provide access to MID, CIR and ESP data stored in CRRS to the EBCGA for the purpose of conducting risk analyses and vulnerability assessments.

RP Specific ID	REQ-01.05 (BR)
Description	The CRRS shall provide access to the duly authorised staff of the European Border and Coast Guard Agency to the ESP, CIR, and MID data stored into CRRS for the purpose of carrying out risk analyses, for monitoring migratory flows towards and within the Union as well as carrying out vulnerability assessments to assess the capacity and readiness of Member States to face upcoming challenges, including present and future threats and challenges at the external borders.
Status	Proposed
Source	01: 02019R0817-20210803 Interoperability Visa Borders REG v1 02: 02019R0818-20210803 Interoperability LEA Immigration REG v1
Rationale	01: Art. 66(4) 02: Art. 62(4)
Keywords	security; access; consult; risk; analyses; vulnerability; assessments; FRONTEX
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.1.6 REQ-01.06 (BR): The CRRS shall enable the duly authorised staff of Europol to have access to data stored in the CRRS related to the CIR and the MID.

RP Specific ID	REQ-01.06 (BR)
Requirement ID	AIC-REQ-00017

RP Specific ID	REQ-01.06 (BR)
Name	The CRRS shall enable the duly authorised staff of Europol to have access to data stored in the CRRS related to the CIR and the MID.
Description	The CRRS shall enable the duly authorised staff of Europol to have access to data stored in the CRRS related to the CIR and the MID, for conducting strategic, thematic and operational analyses.
Status	Proposed
Source	01: 02019R0817-20210803 Interoperability Visa Borders REG v1 02: 02019R0818-20210803 Interoperability LEA Immigration REG v1
Rationale	01: Art. 66(5) 02: Art. 62(5)
Keywords	reporting; ad-hoc; Europol; access; CIR; MID; strategic; operational
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.1.7 REQ-01.07 (BR): The CRRS shall allow the duly authorised staff of the European Border and Coast Guard Agency to access VIS data stored in the CRRS.

RP Specific ID	REQ-01.07 (BR)
Requirement ID	AIC-REQ-00014
Name	The CRRS shall allow the duly authorised staff of the European Border and Coast Guard Agency to access VIS data stored in the CRRS.
Description	The CRRS shall allow the duly authorised staff of the European Border and Coast Guard Agency to have access to statistics related to the following VIS data stored in the CRRS, for the purpose of conducting risk analysis and vulnerability assessments: (a)status information; (b) the authority with which the application has been lodged, including its location; (c) sex, age and nationality or nationalities of the applicant; (d) country and city of residence of the applicant, only as regards visas; (e) current occupation (job group) of the applicant, only as regards visas; (f) the Member States of first entry and destination, only as regards visas; (g) date and place of the application and the decision concerning the application (issued, withdrawn, refused, annulled, revoked, renewed or extended); (h) the type of document applied for or issued, i.e., whether airport transit visa, uniform or limited territorial validity visa, long-stay visa or residence permit; (i) the type of the travel document and the country which issued the travel document, only as regards visas; (j) the decision concerning the application and, in the case of refusal, withdrawal, annulment or revocation, the grounds indicated for that decision; (k) hits resulting from queries of EU information systems, Europol data or Interpol databases pursuant to Article 9a or 22b, differentiated by system or database, or hits against the specific risk indicators pursuant to Article 9j, and hits where, after manual verification pursuant to Article 9c, 9d, 9e or 22b the applicant's personal data was confirmed as corresponding to the data present in one of the information systems or databases queried; (l) decisions to refuse a visa, long-stay visa or residence permit which are correlated to a manually verified and confirmed hit in one of the information systems or databases queried or to a hit against the specific risk indicators; (m) the competent authority, including its location, which decided on the

RP Specific ID	REQ-01.07 (BR)
	application and the date of the decision, only as regards visas; (n) the cases in which the same applicant applied for a visa from more than one visa authority, indicating those visa authorities, their location and the dates of the decisions; (o) the main purposes of the journey, only as regards visas; (p) visa applications processed in representation pursuant to Article 8 of Regulation (EC) No 810/2009; (q) the data entered in respect of any document withdrawn, annulled, revoked, renewed or extended, as applicable; (r) the expiry date of the long-stay visa or residence permit; (s) the number of persons exempt from the requirement to give fingerprints pursuant to Article 13(7) of Regulation (EC) No 810/2009; (t) the cases in which the data referred to in point (6) of Article 9 could not be provided, in accordance with Article 8(5); (u) the cases in which the data referred to in point (6) of Article 9 was not required to be provided for legal reasons, in accordance with Article 8(5); (v) the cases in which a person who could not provide the data referred to in point (6) of Article 9 was refused a visa, in accordance with Article 8(5); (w) links to the previous application file on that applicant as well as links of the application files of the persons travelling together, only as regards visas.
Status	Proposed
Source	01: 32021R1134-20210713 VIS REG
Rationale	01: Art. 45a(1)
Keywords	risk analysis; vulnerability assessment; FRONTEX; VIS
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager

RP Specific ID	REQ-01.07 (BR)
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.1.8 REQ-01.08 (BR): The CRRS shall allow the duly authorised staff of the competent authorities of MS, the COM, eu-LISA, the European Asylum Support Office and the FRONTEX, including the ETIAS CU, to have access to consult VIS data stored in the CRRS.

RP Specific ID	REQ-01.08 (BR)
Requirement ID	AIC-REQ-00010
Name	The CRRS shall allow the duly authorised staff of the competent authorities of MS, the COM, eu-LISA, the European Asylum Support Office and the FRONTEX, including the ETIAS CU, to have access to consult VIS data stored in the CRRS.
Description	The CRRS shall allow the duly authorised staff of the competent authorities of Member States, the Commission, eu-LISA, the European Asylum Support Office and the European Border and Coast Guard Agency, including the ETIAS Central Unit, to have access to consult the following VIS data stored in the CRRS, without allowing for individual identification: (a) status information; (b) the authority with which the application has been lodged, including its location; (c) sex, age and nationality or nationalities of the applicant; (d) country and city of residence of the applicant, only as regards visas; (e) current occupation (job group) of the applicant, only as regards visas; (f) the Member States of first entry and destination, only as regards visas; (g) date and place of the application and the decision concerning the application (issued, withdrawn, refused, annulled, revoked, renewed or extended); (h) the type of document applied for or issued, i.e., whether airport transit visa, uniform or limited territorial validity visa, long-stay visa or residence permit; (i) the type of the travel document and the country which issued the

RP Specific ID	REQ-01.08 (BR)
	travel document, only as regards visas; (j) the decision concerning the application and, in the case of refusal, withdrawal, annulment or revocation, the grounds indicated for that decision; (k) hits resulting from queries of EU information systems (SIS, EES, ETIAS including the ETIAS watchlist, Eurodac, ECRIS-TCN), Europol data, Interpol databases (Stolen and Lost Travel Document database and Interpol Travel Documents Associated with Notices database) differentiated by system or database, hits against the VIS risk indicators, and hits where, after manual verification the applicant's personal data was confirmed as corresponding to the data present in one of the information systems or databases queried; (l) decisions to refuse a visa, long-stay visa or residence permit which are correlated to a manually verified and confirmed hit in one of the information systems or databases queried or to a hit against the specific risk indicators; (m) the competent authority, including its location, which decided on the application and the date of the decision, only as regards visas; (n) the cases in which the same applicant applied for a visa from more than one visa authority, indicating those visa authorities, their location and the dates of the decisions; (o) the main purposes of the journey, only as regards visas; (p) visa applications processed in representation (a Member State examining and deciding on applications on behalf of another Member State); (q) the data entered in respect of any document withdrawn, annulled, revoked, renewed or extended, as applicable; (r) the expiry date of the long-stay visa or residence permit; (s) the number of persons exempt from the requirement to give fingerprints; (t) the cases in which the fingerprints of the applicant could factually not provided; (u) the cases in which the fingerprints of the applicant were not required to be provided for legal reasons; (v) the cases in which a person who could not provide fingerprints was refused a visa; (w) links to the previous application file on that applicant as well as links of the application files of the persons travelling together, only as regards visas.
Status	Proposed
Source	01: 32021R1134-20210713 VIS REG

RP Specific ID	REQ-01.08 (BR)
Rationale	01: Art. 45a(2)
Keywords	data; reporting; statistics; VIS; competent authorities; COM; eu-LISA, EASO; FRONTEX; individual identification
Priority	Must have
Target / Sub target	AI in CRRS\Intake
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.2 Stakeholder Requirements (RP.01)

9.1.2.1 REQ-01.00.01 (SHR): The CRRS reporting tool shall enable cross-system statistical data and analytical reporting where provided for by Union law.

RP Specific ID	REQ-01.00.01 (SHR)
Requirement ID	AIC-REQ-00073
Name	The CRRS reporting tool shall enable cross-system statistical data and analytical reporting where provided for by Union law.
Description	The CRRS reporting tool shall enable the provision of cross-system statistical data and analytical reporting for policy, operational and data quality purposes where provided for by Union Law.
Status	Proposed
Source	01: 32021R2223-20210930 Operation of CRRS supplementing

RP Specific ID	REQ-01.00.01 (SHR)
	2019/817_DA 01: 32021R2222-20210930 Operation of CRRS supplementing 2019/818 DA
Rationale	01: Art. 3(4) 02: Art. 3(4)
Keywords	monitoring; functioning; statistics; reports; vulnerability; assessment; risk; analysis; FRONTEX; EES; VIS; ETIAS; Eurodac; ECRIS-TCN; SIS
Priority	
Target / Sub target	CRRS/Uptake
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.2.2 REQ-01.00.02 (SHR): The CRRS shall obtain from EU information systems and IO components read-only copies of the necessary data.

RP Specific ID	REQ-01.00.02 (SHR)
Requirement ID	AIC-REQ-00074
Name	The CRRS shall obtain from EU information systems and IO components read-only copies of the necessary data.
Description	The CRRS shall obtain from the EU information systems and interoperability components read-only copies of the necessary data to produce the statistics and reports regarding EES, VIS, ETIAS, Eurodac* SIS and ECRIS-TCN, and regarding ESP, CIR and MID. The CRRS shall obtain the data on a regular basis and at least daily, through a one-way transfer.

RP Specific ID	REQ-01.00.02 (SHR)
	* Subject to approval of proposals for amendments to Regulation (EU) 2013/ 603 [Eurodac] leading to the adoption of amending regulations.
Status	Proposed
Source	01: 32021R2223-20210930 Operation of CRRS supplementing 2019/817_DA 01: 32021R2222-20210930 Operation of CRRS supplementing 2019/818 DA
Rationale	01: Art. 4 02: Art. 4
Keywords	copy; one way; read-only; extraction; EES; VIS; ETIAS; EURODAC; ECRIS-TCN; SIS; ESP; MID; CIR; FRONTEX
Priority	
Target / Sub target	CRRS/Intake
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.2.3 REQ-01.01.01 (SHR): The CRRS shall allow to define cluster characteristics

RP Specific ID	REQ-01.01.01 (SHR)
Requirement ID	AIC-REQ-00078
Name	The CRRS shall allow to define cluster characteristics (i.e., which information is required to be displayed within the extracted groups of travellers).

RP Specific ID	REQ-01.01.01 (SHR)
Description	The CRRS shall provide analytical capabilities to allow to render and merge data from all CBS (EES, ETIAS, VIS, EURODAC, ECRIS TCN, VIS and SIS II) to set up different scenarios, using risk-levels and additional risk threshold as agreed among eu-LISA experts and stakeholders to display the extracted groups of travellers.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	CBS; clusters; information; display; groups; travellers
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.2.4 REQ-01.01.02 (SHR): The CRRS shall allow to define the risk-level severity, upon agreement among eu-LISA's stakeholders.

RP Specific ID	REQ-01.01.02 (SHR)
Requirement ID	AIC-REQ-00077
Name	The CRRS shall allow to define the risk-level severity, upon agreement among eu-LISA's stakeholders.

Specific	REQ-01.01.02 (SHR)
Description	The CRRS shall provide analytical capabilities to allow to define different levels of risk severity from the historical data available and upon agreement among eu-LISA's stakeholders. Levels can be classified into either level type or Indicator type: i) Level Type: very-high risk, high risk, medium risk, low risk levels ii) Indicator Type: security-risk, illegal immigration-risk, high epidemic-risk indicators.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	risk; level; severity; Indicator; data
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.2.5 REQ-01.01.03 (SHR): The CRRS shall allow to determine the appropriate number of groups of travellers (i.e., clusters).

RP Specific ID	REQ-01.01.03 (SHR)
Requirement ID	AIC-REQ-00076

RP Specific ID	REQ-01.01.03 (SHR)
Name	The CRRS shall allow to determine the appropriate number of groups of travellers (i.e., clusters).
Description	The CRRS shall provide analytical capabilities to allow to determine the appropriate number of groups of travellers that will form the clusters, which is obtained from the historical data available. This will be done by either under common agreement among eu-LISA's stakeholders (i.e. setting up a predefined number of clusters), or through an analysis that will indicate the number of groups of travellers that present similarities among them.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	number; group; travellers; similarities; predefined; clusters; data; agreement; analysis
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, and the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.2.6 REQ-01.02.01 (SHR): The CRRS shall provide statistical reports to the duly authorised staff of the competent authorities on the records kept in the EES.

RP Specific ID	REQ-01.02.01 (SHR)
Requirement ID	AIC-REQ-00032
Name	The CRRS shall provide statistical reports to the duly authorised staff of the competent authorities on the records kept in the EES.
Description	The CRRS shall provide to the duly authorised staff of the competent authorities of Member States, the Commission, the European Border and Coast Guard Agency and eu-LISA reports containing the following statistics on records kept in the EES: a) customisable reports and statistics on entries and exits, refusals of entry and overstays of third country nationals; b) daily statistics on overstayers, third-country nationals who were refused entry, third-country nationals whose authorisation for stay was revoked or extended and third country nationals exempt from the requirement to give fingerprints; c) customisable reports and statistics on data quality and regular statistics for ensuring the monitoring by eu-LISA of the development and the functioning of the EES.
Status	Proposed
Source	01: 32021R2223-20210930 Operation of CRRS supplementing 2019/817_DA
Rationale	01: Art. 2(5)(a)
Keywords	EES; customisable; statistics; overstayers; refused; data quality; monitoring; FRONTEX; member; states; commission; parliament; council
Priority	Should Have
Target / Sub target	CRRS/Uptake
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager

RP Specific ID	REQ-01.02.01 (SHR)
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, and the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.1.2.7 REQ-01.04.01 (SHR): The AI technology shall have access to data related to the applicants' provided in the application forms and data related to the TAs decisions.

RP Specific ID	REQ-01.04.01 (SHR)
Requirement ID	AIC-REQ-00039
Name	The AI technology shall have access to data related to the applicants' provided in the application forms and data related to the TAs decisions.
Description	The AI technology shall have access to the following data in the ETIAS IS though the CRRS, without allowing for individual identification: (a) application status information; (b) nationalities, sex and year of birth of the applicant; (c) the country of residence; (d) education (primary, secondary, higher or none); (e) current occupation (job group); (f) the type of the travel document and three-letter code of the issuing country; (g) the type of TA and, for a TA with TV on humanitarian grounds, a reference to the MSs issuing the LTV TA; (h) the validity period of the TA; and (i) the grounds for refusing, revoking or annulling a TA.
Status	Proposed
Source	01: ANNEX 8.2 ETIAS Business Use Case Model
Rationale	01: Sec. 11.1.3.7(1)(2)
Keywords	anonymised; data; identity; TA; decision; CRRS

RP Specific ID	REQ-01.04.01 (SHR)
Priority	Should Have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.2 RP.02 Review and validate identified risk profile.

9.2.1 Business Requirements (RP.02)

9.2.1.1 REQ-02.01 (BR): The ETIAS screening rules shall be an algorithm enabling profiling through comparison of data recorded in ETIAS CS with specific risk indicators.

RP Specific ID	REQ-02.01 (BR)
Requirement ID	AIC-REQ-00001
Name	The ETIAS screening rules shall be an algorithm enabling profiling through comparison of data recorded in ETIAS CS with specific risk indicators.
Description	The ETIAS screening rules shall be an algorithm enabling profiling through the comparison of the data recorded in an application file of the ETIAS CS with specific risk indicators established by the ETIAS CU pointing to security, illegal immigration or high epidemic risks. The ETIAS CU shall register the ETIAS screening rules in the ETIAS CS.
Status	Proposed

RP Specific ID	REQ-02.01 (BR)
Source	01: 02018R1240-20210803 ETIAS REG
Rationale	01: Art. 33(1)
Keywords	screening rules; algorithm; risk; indicator
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.2.1.2 REQ-02.02 (BR): The ETIAS IS shall allow the ETIAS CU to establish a set of specific risk indicators based on the specific risks related to security or illegal immigration or a high epidemic risk.

RP Specific ID	REQ-02.02 (BR)
Requirement ID	AIC-REQ-00006
Name	The ETIAS IS shall allow the ETIAS CU to establish a set of specific risk indicators based on the specific risks related to security or illegal immigration or a high epidemic risk.
Description	Based on the specific risks related to security or illegal immigration or a high epidemic risk, The ETIAS IS shall allow the ETIAS CU to establish a set of specific risk indicators consisting of a combination of data including one or more of the following: (a) age range, sex, nationality; (b) country and city of residence;

RP Specific ID	REQ-02.02 (BR)
	(c) level of education (primary, secondary, higher or none); (d) current occupation (job group). The specific risk indicators shall be targeted and proportionate. They shall in no circumstances be based solely on a person's sex or age. They shall in no circumstances be based on information revealing a person's colour, race, ethnic or social origin, genetic features, language, political or any other opinion, religion or philosophical belief, trade union membership, membership of a national minority, property, birth, disability or sexual orientation. The specific risk indicators shall be defined, established, assessed ex-ante, implemented, evaluated ex-post, revised and deleted by the ETIAS CU.
Status	Proposed
Source	01: 02018R1240-20210803 ETIAS REG
Rationale	01: Art. 33(4), 33(5), 33(6)
Keywords	risk; indicators; ex-ante; ex-post; combination of data
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.2.1.3 REQ-02.03 (BR): The CRRS shall provide statistical reports to the ETIAS Central Unit supporting them in defining specific risk indicators for the VIS.

RP Specific ID	REQ-02.03 (BR)
Requirement ID	AIC-REQ-00015
Name	The CRRS shall provide statistical reports to the ETIAS Central Unit supporting them in defining specific risk indicators for the VIS.
Description	The CRRS shall provide statistical reports to the ETIAS Central Unit supporting them in defining specific risk indicators for the VIS. The risk indicators shall be based on a combination of the following data: - age range, sex, nationality; - country and city of residence; - the Member States of destination; - the Member State of first entry; - purpose of travel; - current occupation (job group).
Status	Proposed
Source	01: 32021R1134-20210713 VIS REG
Rationale	01: Art. 9j(4)
Keywords	data management; metadata management; ETIAS; VIS; travel; application; risk; age; sex; nationality; country; city; destination; first entry; travel; purpose; occupation
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.

RP Specific ID	REQ-02.03 (BR)
Sector	(BRMS) Business Relations Management Sector

9.2.2 Stakeholder Requirements (RP.02)

9.2.2.1 REQ-02.00.01 (SHR): The CRRS shall provide statistics and reports that allow ETIAS Central Unit to continuously establish and identify risk indicators.

RP Specific ID	REQ-02.00.01 (SHR)
Requirement ID	AIC-REQ-00075
Name	The CRRS shall provide statistics and reports that allow ETIAS Central Unit to continuously establish and identify risk indicators.
Description	The CRRS statistics and reports about EES and ETIAS shall allow ETIAS Central Unit to establish the specific risk indicators used as part of the screening process, related to illegal immigration or epidemic outbreaks. This CRRS statistics and reports shall allow identification of sets of characteristics corresponding to specific groups of travellers[90] associated with security or illegal immigration or high epidemic risks. In order to ensure that risks may be defined on the basis of CRRS statistics about EES and ETIAS, it is necessary to ensure that such elements are monitored and collected by the ETIAS and the EES in a manner that allows for the continuous identification of sets of characteristics corresponding to specific groups of travellers associated with overstaying, refusal of entry or refusal of travel authorisation: total number of cases associated with such practises, identify specific groups of travellers for which such practises occur at abnormal rates, identify the sets of characteristics corresponding to the specific groups of travellers in question as well as any correlation with information collected through the application files.
Status	Proposed
Source	01: C(2021)4981-20211123 ETIAS Risks DA
Rationale	01: Art. 3

[90] In compliance with the principles established in the EU fundamental rights, the sets of characteristics of specific group of travellers shall not lead to direct or indirect discrimination

RP Specific ID	REQ-02.00.01 (SHR)
Keywords	EES; ETIAS; risk indicators; screening; illegal immigration; epidemic outbreaks; FRONTEX
Priority	
Target / Sub target	CRRS/Uptake
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.2.2.2 REQ-02.01.01 (SHR): The CRRS shall ensure enough historical data to perform review and validation of identified groups of travellers of risk.

RP Specific ID	REQ-02.01.01 (SHR)
Requirement ID	AIC-REQ-00080
Name	The CRRS shall ensure enough historical data to perform review and validation of identified groups of travellers of risk.
Description	The CRRS will execute its analytical capabilities to review and validate identified groups of travellers of risk when enough data is available from all CBS (EES, ETIAS, VIS, EURODAC, ECRIS TCN, VIS and SIS II).
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	CBS; precondition; historical; data; analysis; processes

RP Specific ID	REQ-02.01.01 (SHR)
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.2.2.3 REQ-02.01.02 (SHR): The CRRS shall allow to define anomaly detection levels corresponding to the groups of travellers.

RP Specific ID	REQ-02.01.02 (SHR)
Requirement ID	AIC-REQ-00079
Name	The CRRS shall allow to define anomaly detection levels corresponding to the groups of travellers.
Description	The CRRS shall provide analytical capabilities to allow to render and merge data from all CBS (EES, ETIAS, VIS, EURODAC, ECRIS TCN, VIS and SIS II) and set up different anomaly detection levels as agreed among eu-LISA experts and stakeholders, to display either "normal" or "suspicious" (or any other level agreed) anomaly levels from a particular set of incoming groups of travellers.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A

RP Specific ID	REQ-02.01.02 (SHR)
Keywords	CBS; information; anomaly; levels; detection; groups; travellers
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.2.2.4 REQ-02.01.03 (SHR): The ETIAS CS shall enable establishing a new risk profile by laying down a standard format for each risk in the ETIAS framework.

RP Specific ID	REQ-02.01.03 (SHR)
Requirement ID	AIC-REQ-00055
Name	The ETIAS CS shall enable establishing a new risk profile by laying down a standard format for each risk in the ETIAS framework.
Description	The ETIAS Central Unit shall specify the security, illegal immigration or high epidemic risks based on the analysis of (i) statistical data on overstaying, refusal of entry and refusal of travel authorisation; (ii) specific security risks or threats (iii) abnormal rates of overstaying and refusals of entry; (iv) specific high epidemic risks provided by the MS, the ECDC and the WHO by laying down in a standard format for each risk: (a) the applicable risk category or categories (illegal immigration, security or high epidemic risk); (b) a name in free text, which shall be unique within the same risk

RP Specific ID	REQ-02.01.03 (SHR)
	category; (c) a description of the risk in free text, including the sets of characteristics of specific group of travellers[91] associated with the risk and the size of the specific group considered; (d) guidance for ETIAS NU to consider when assessing a hit related to the risk; (e) the detailed explanation of and objective justification for the application of different rules regarding the three months period considered for the assessment; (f) the unique reference number of the related risk analysis conducted, specifying where relevant the MS at the origin of the information or analysis on which the risk analysis is based; (g) and for the risks identified in accordance with analysis (i) to (iii) above, the rate and absolute number of overstayers, refusals of entry or refusals of travel authorisations considered normal, and the deviation from the normal rate and absolute number signalling an abnormal condition. The ETIAS Central Unit shall establish a list of all the risks specified in accordance with the above ("the list").
Status	Proposed
Source	01: C(2021)4981-20211123 ETIAS Risks DA
Rationale	01: Art. 3(1)(2)
Keywords	new; risk profile; analysis; overstay; security; illegal immigration; high epidemic; refusal of entry; standard format
Priority	Must Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)

[91] In compliance with the principles established in the EU fundamental rights, the sets of characteristics of specific group of travellers shall not lead to direct or indirect discrimination

RP Specific ID	REQ-02.01.03 (SHR)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.2.2.5 REQ-02.01.04 (SHR): The ETIAS CS shall enable establishing a new risk profile by laying down a standard format for each risk using VIS data.

RP Specific ID	REQ-02.01.04 (SHR)
Requirement ID	AIC-REQ-00054
Name	The ETIAS CS shall enable establishing a new risk profile by laying down a standard format for each risk using VIS data.
Description	The ETIAS Central Unit shall specify the security, illegal immigration or high epidemic risks based on the analysis of (i) statistical data on overstaying, refusal of entry and refusal of visa applications; (ii) specific security risks or threats (iii) abnormal rates of overstaying and refusals of entry; (iv) specific high epidemic risks provided by the MS, the ECDC and the WHO by laying down in a standard format for each risk: (a) the applicable risk category or categories (illegal immigration, security or high epidemic risk); (b) a name in free text, which shall be unique within the same risk category; (c) a description of the risk in free text, including the sets of characteristics of specific group of visa holders[92] associated with the risk and the size of the specific group considered; (d) guidance for ETIAS NU to consider when assessing a hit related to the risk;

[92] In compliance with the principles established in the EU fundamental rights, the sets of characteristics of specific group of travellers shall not lead to direct or indirect discrimination

RP Specific ID	REQ-02.01.04 (SHR)
	(e) the detailed explanation of and objective justification for the application of different rules regarding the three months period considered for the assessment; (f) the unique reference number of the related risk analysis conducted, specifying where relevant the MS at the origin of the information or analysis on which the risk analysis is based; (g) and for the risks identified in accordance with analysis (i) to (iii) above, the rate and absolute number of overstayers, refusals of entry or refusals of visa applications considered normal, and the deviation from the normal rate and absolute number signalling an abnormal condition. The ETIAS Central Unit shall establish a list of all the risks specified in accordance with the above ("the list").
Status	Postponed
Source	01: X(2021)-20211217 Definition of risks DA
Rationale	01: Art. 3(1)(2)
Keywords	new; risk profile; analysis; overstay; security; illegal immigration; high epidemic; refusal of entry; standard format
Priority	Must Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.2.2.6 REQ-02.01.05 (SHR): The ETIAS CS shall allow to assess normal and abnormal rates by comparing similar periods of time and defining thresholds, including acceptable error margins.

RP Specific ID	REQ-02.01.05 (SHR)
Requirement ID	AIC-REQ-00061
Name	The ETIAS CS shall allow to assess normal and abnormal rates by comparing similar periods of time and defining thresholds, including acceptable error margins.
Description	For the purposes of assessing normal and abnormal rates, the ETIAS Central Unit and Member States shall apply the following rules: 1. the period considered when assessing whether a rate and absolute number is normal or abnormal shall be three months. This period shall be compared to a similar period of time from the previous year or years or, where necessary due to objective reasons relating to the nature of the event, from any other year or years; 2. the ETIAS Central Unit and Member States shall, for each risk identified determine specific thresholds for considering a rate and absolute number normal for the period in question; 3. where the rate and absolute number is above the thresholds determined, the rate and absolute number shall be considered abnormal, representing a heightened risk. The rate and absolute number in question shall not be considered abnormal if it can be explained by other factors, including: (a) periodical variations in line with historical observations, including seasonality; (b) the modification of border management processes or systems, including due to new legislation or technology; (c) stricter border control checks, including ones due to joint or national operations; (d) singular events, including international sports events or national elections; 4. for the purpose of point 3, the ETIAS Central Unit and Member States shall specify acceptable error margins.
Status	Postponed
Source	01: X(2022)-20220304 Specification of risks IA 02: X(2021)-20211217 Specification of risks according to Art. 9j3 IA
Rationale	01: Art. 2 02: Art. 2
Keywords	risk profiles; threshold; assess; error margins; 3 months; rate; absolute number; error margins

RP Specific ID	REQ-02.01.05 (SHR)
Priority	Should Have
Target / Sub target	
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.2.2.7 REQ-02.02.01 (SHR): The AI technology shall provide ex-ante assessments.

RP Specific ID	REQ-02.02.01 (SHR)
Requirement ID	AIC-REQ-00041
Name	The AI technology shall provide ex-ante assessments.
Description	The AI technology shall provide ex-ante assessment analysis of identified risk profiles, including the matching rate percentage and the result of the calculation of inclusivity and reliability. The system shall allow to introduce and save comments to the assessments. The system shall allow ex-ante assessment to be exported. The different analyses shall not contain any personal data.
Status	Proposed
Source	01: Annex 8.2 ETIAS Business Use Case Model 02: ETIAS Risk Screening BRD
Rationale	01: Sec. 9.5.10 02: Art. 8
Keywords	ex-ante assessment; risk profile; data protection

RP Specific ID	REQ-02.02.01 (SHR)
Priority	Should Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.2.2.8 REQ-02.02.02 (SHR): The AI technology shall perform comparative analysis of risk profiles performance (ex-ante).

RP Specific ID	REQ-02.02.02 (SHR)
Requirement ID	AIC-REQ-00042
Name	The AI technology shall perform comparative analysis of risk profiles performance (ex-ante).
Description	The AI technology shall perform comparative analysis of several risk profiles based on their ex-ante assessment results. The AI technology shall identify patterns, trends and behaviours. The system shall allow outputs to be exported.
Status	Proposed
Source	01: Annex 8.2 ETIAS Business Use Case Model 02: ETIAS Risk Screening BRD
Rationale	01: Sec. 9.5.12 02: Sec. 8.2
Keywords	ex-ante assessment; risk profile; comparison report

RP Specific ID	REQ-02.02.02 (SHR)
Priority	Should Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.2.2.9 REQ-02.02.03 (SHR): The AI technology shall apply evaluation criteria following the ETIAS CS tool used to assess the impact of new risk indicators (ex-ante).

RP Specific ID	REQ-02.02.03 (SHR)
Requirement ID	AIC-REQ-00038
Name	The AI technology shall apply evaluation criteria following the ETIAS CS tool used to assess the impact of new risk indicators (ex-ante).
Description	The AI technology shall use the evaluation criteria's existing in the ETIAS CS tool (ex-ante assessment) as means to identifying potential new risks based on their impact.
Status	Proposed
Source	01: Notes on AG/PMB support
Rationale	01: Notes on AG/PMB support - 25-03-2019
Keywords	risk; evaluate; impact; ex-ante
Priority	Should Have

RP Specific ID	REQ-02.02.03 (SHR)
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.3 RP.03 Analyse risk profiles correlations.

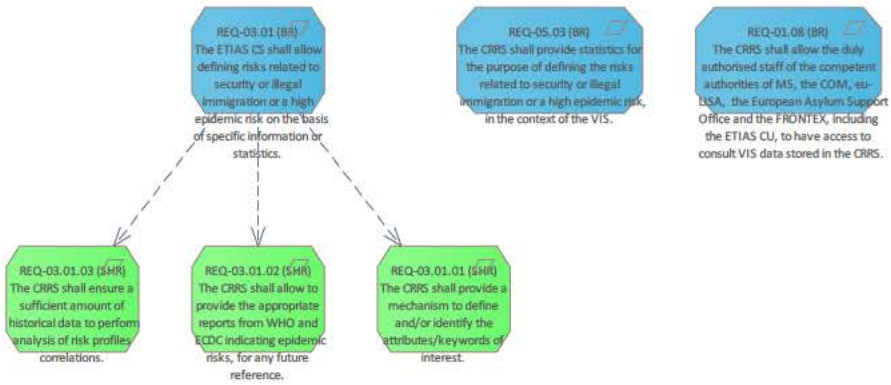


Figure 53. RP.03 Analyse risk profiles correlations.

9.3.1 Business Requirements (RP.03)

9.3.1.1 REQ-03.01 (BR): The ETIAS CS shall allow defining risks related to security or illegal immigration or a high epidemic risk on the basis of specific information or statistics.

RP Specific ID	REQ-03.01 (BR)
Requirement ID	AIC-REQ-00002
Name	The ETIAS CS shall allow defining risks related to security or illegal immigration or a high epidemic risk based on specific information or statistics.
Description	The ETIAS CS shall allow to define the risks related to security or illegal immigration or a high epidemic risk based on: (a) statistics generated by the EES indicating abnormal rates of overstaying and refusals of entry for a specific group of travellers; (b) statistics generated by the ETIAS, aiming the use of data for reporting and statistics, indicating abnormal rates of refusals of TAs due to a security, illegal immigration or high epidemic risk associated with a specific group of travellers; (c) statistics generated by the ETIAS and the EES indicating correlations between information collected through the application form and overstaying by travellers or refusals of entry; (d) information substantiated by factual and evidence-based elements provided by MSs concerning specific security risk indicators or threats identified by that MS; (e) information substantiated by factual and evidence-based elements provided by MSs concerning abnormal rates of overstaying and refusals of entry for a specific group of travellers for that MS; (f) information concerning specific high epidemic risks provided by MSs as well as epidemiological surveillance information and risk assessments provided by the ECDC and disease outbreaks reported by the WHO. The system shall allow the specific risks to be reviewed at least every 6 months.
Status	Proposed
Source	01: 02018R1240-20210803 ETIAS REG
Rationale	01: Art. 33(2), 33(3), 84
Keywords	risk; security; illegal; immigration; epidemic; screening rules

RP Specific ID	REQ-03.01 (BR)
Priority	Must have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.3.2 Stakeholder Requirements (RP.03)

9.3.2.1 REQ-03.01.01 (SHR): The CRRS shall provide a mechanism to define and/or identify the attributes/keywords of interest.

RP Specific ID	REQ-03.01.01 (SHR)
Requirement ID	AIC-REQ-00083
Name	The CRRS shall provide a mechanism to define and/or identify the attributes/keywords of interest.
Description	The CRRS shall provide analytical capabilities to allow to define and extract attributes/keywords from WHO/ECDC reports, to identify and highlight correlations amongst risk indicators.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	attributes; keywords; define; extracted; correlations; risk indicators

RP Specific ID	REQ-03.01.01 (SHR)
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.3.2.2 REQ-03.01.02 (SHR): The CRRS shall allow to provide the appropriate reports from WHO and ECDC indicating epidemic risks, for any future reference.

RP Specific ID	REQ-03.01.02 (SHR)
Requirement ID	AIC-REQ-00082
Name	The CRRS shall allow to provide the appropriate reports from WHO and ECDC indicating epidemic risks, for any future reference.
Description	The CRRS shall provide analytical capabilities to allow to render information from WHO/ECDC reports, for the process of identifying and highlighting correlations amongst risk indicators.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	WHO; ECDC; reports; rendering; correlations; risk indicators
Priority	Must have

RP Specific ID	REQ-03.01.02 (SHR)
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, and the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.3.2.3 REQ-03.01.03 (SHR): The CRRS shall ensure enough historical data to perform analysis of risk profiles correlations.

RP Specific ID	REQ-03.01.03 (SHR)
Requirement ID	AIC-REQ-00081
Name	The CRRS shall ensure enough historical data to perform analysis of risk profiles correlations.
Description	The CRRS will execute its analytical capabilities to analyse risk profiles correlations when enough data is available from all CBS (EES, ETIAS, VIS, EURODAC, ECRIS TCN, VIS and SIS II).
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	CBS; precondition; historical; data; correlation analysis
Priority	Must have
Target / Sub target	CRRS

RP Specific ID	REQ-03.01.03 (SHR)
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, and the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.4 RP.04 Automatic processing of textual information from various sources

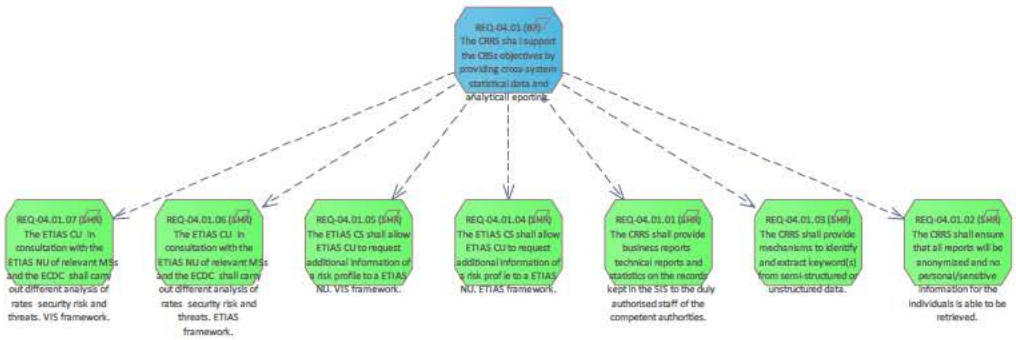


Figure 54. RP.04 Automatic processing of textual information from various sources.

9.4.1 Business Requirements (RP.04)

9.4.1.1 REQ-04.01 (BR): The CRRS shall support the CBSs objectives by providing cross-system statistical data and analytical reporting.

RP Specific ID	REQ-04.01 (BR)
Requirement ID	AIC-REQ-00018
Name	The CRRS shall support the CBSs objectives by providing cross-system statistical data and analytical reporting.
Description	The CRRS shall support the objectives of the EES, VIS, ETIAS, Eurodac, ECRIS-TCN and SIS by providing cross-system statistical data and

RP Specific ID	REQ-04.01 (BR)
	analytical reporting for policy, operational and data quality purposes.
Status	Proposed
Source	01: 02019R0817-20210803 Interoperability Visa Borders REG v1 02: 02019R0818-20210803 Interoperability LEA Immigration REG v1
Rationale	01: Art. 39(1) 02: Art. 39(1)
Keywords	central repository; components; statistical data; eu information systems; FRONTEX; EES; VIS; ETIAS; Eurodac; ECRIS-TCN; SIS
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.4.2 Stakeholder Requirements (RP.04)

9.4.2.1 REQ-04.01.01 (SHR): The CRRS shall provide business reports, technical reports and statistics on the records kept in the SIS to the duly authorised staff of the competent authorities.

RP Specific ID	REQ-04.01.01 (SHR)
Requirement ID	AIC-REQ-00035
Name	The CRRS shall provide business reports, technical reports and statistics on the records kept in the SIS to the duly authorised staff of the competent authorities.

RP Specific ID	REQ-04.01.01 (SHR)
Description	The CRRS shall provide reports from SIS data in the field of police cooperation and judicial cooperation in criminal matters, containing the following statistics on the records kept in the SIS to the duly authorised staff of the competent authorities: a) Daily, monthly and annual statistics showing the number of records per category of alerts, both for each Member State and in aggregate, b) Annual reports on the number of hits per category of alert, how many times SIS was searched and accessed for the purpose of entering, updating or deleting an alert, both for each Member State and in aggregate, c) at the request of the Commission, additional specific statistical reports, either on a regular or ad hoc basis, on the performance of SIS, the use of SIS and on the exchange of supplementary information, d) at the request of the European Border and Coast Guard Agency, additional specific statistical reports, either on a regular or ad hoc basis, for the purpose of conducting risk analyses and vulnerability assessments e) Reports and statistics for the purposes of technical maintenance, reporting, data quality reporting and statistics, f) data quality reports.
Status	Proposed
Source	01: 32021R2222-20210930 Operation of CRRS supplementing 2019/818 DA
Rationale	01: Art. 2(5)(a)
Keywords	provide; SIS II; reports; statistics; records; kept; FRONTEX; data quality
Priority	Must Have
Target / Sub target	CRRS/Uptake CRRS/Platform Management
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised

RP Specific ID	REQ-04.01.01 (SHR)
	and validated/tested, and the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.4.2.2 REQ-04.01.02 (SHR): The CRRS shall ensure that all reports will be anonymized and no personal/sensitive information for the individuals is able to be retrieved.

RP Specific ID	REQ-04.01.02 (SHR)
Requirement ID	AIC-REQ-00085
Name	The CRRS shall ensure that all reports will be anonymized and no personal/sensitive information for the individuals is able to be retrieved.
Description	The CRRS shall provide analytical capabilities to allow feature extraction of text report in an anonymous manner, ensuring no personal/sensitive information can be retrieved.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	data; extraction; anonymization
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, and the description of this stakeholder requirement is fulfilled and it becomes 'true'.

RP Specific ID	REQ-04.01.02 (SHR)
Sector	(BRMS) Business Relations Management Sector

9.4.2.3 REQ-04.01.03 (SHR): The CRRS shall provide mechanisms to identify and extract keyword(s) from semi-structured or unstructured data.

RP Specific ID	REQ-04.01.03 (SHR)
Requirement ID	AIC-REQ-00084
Name	The CRRS shall provide mechanisms to identify and extract keyword(s) from semi-structured or unstructured data.
Description	The CRRS shall provide analytical capabilities to allow to identify keywords that will be then extracted via automatic text classification process (e.g., Status Information", "VISA Authority", "Location of Visa Authority", "Nationality", "Member State", "Date of VISA Application", "Place of VISA Application", "Type of VISA", "Type of Travel Document", "Number of VISA Applications" , "Status of VISA Application", "Date of Acceptance", "Date of Refusal", "Purpose of Journey", "Fingerprints", etc.)
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	target; keyword; automatic; classification; process
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager

RP Specific ID	REQ-04.01.03 (SHR)
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.4.2.4 REQ-04.01.04 (SHR): The ETIAS CS shall allow ETIAS CU to request additional information of a risk profile to a ETIAS NU. ETIAS framework.

RP Specific ID	REQ-04.01.04 (SHR)
Requirement ID	AIC-REQ-00057
Name	The ETIAS CS shall allow ETIAS CU to request additional information of a risk profile to a ETIAS NU. ETIAS framework.
Description	The ETIAS Central System (CS) shall allow ETIAS Central Unit (CU) to request analysis of a risk profile from the ETIAS NU. The analysis and any review of it shall include at least: (a) a description of the security risk or threat identified, including the frequency, trends and impacts of past occurrences of incidents, or the risks associated with abnormal rates of overstaying or refusals of entry identified, specifying a unique reference number for each risk or threat; (b) a list of known facts, evidence and intelligence related to the security risk or threat identified or evidence concerning: (i) an abnormal rate or absolute number of overstayers for a specific group of travellers' persons present in that Member State or in a specific area of that Member State compared to other specific groups of travellers' persons present in that Member State or in a specific area of that Member State; or (ii) an abnormal rate or absolute number of refusals of entry issued by that Member State for a specific group of travellers' persons compared to the number of refusals of entry issued by that Member State for other specific groups of travellers' persons; or (c) the sets of characteristics of specific groups of travellers[93] potentially posing or being particularly exposed to the security risk or

[93] In compliance with the principles established in the EU fundamental rights, the sets of characteristics of specific group of travellers shall not lead to direct or indirect discrimination

RP Specific ID	REQ-04.01.04 (SHR)
	threat identified or those associated with overstaying or refusal of entry potentially posing or being particularly exposed to the risk identified, considering the following data: (i) age range, sex, nationality; (ii) country and city of residence; (iii) level of education (primary, secondary, higher or none); (iv) current occupation (job group). Where one or more of the elements referred to in point (a) are not available, a justification shall be provided by ETIAS NU to ETIAS CU. Whenever a request is saved, the system shall update the risk indicator and send a notification to the ETIAS NU.
Status	Proposed
Source	01: Annex 8.2 ETIAS Business Use Case Model 02: ETIAS Risk Screening BRD 03: C(2021)4981-20211123 ETIAS Risks DA
Rationale	01: Sec. 9.5.14 02: Sec. 4.1(3.1.5) 03: Art.4(1)(a)(b)(c), Art.4(2), Art. 5(1)(a)(b)(c), Art.5(2)
Keywords	request additional information; risk profile; notification
Priority	Should Have
Target / Sub target	ETIAS/Screening Rule
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.4.2.5 REQ-04.01.05 (SHR): The ETIAS CS shall allow ETIAS CU to request additional information of a risk profile to a ETIAS NU. VIS framework.

RP Specific ID	REQ-04.01.05 (SHR)
Requirement ID	AIC-REQ-00062
Name	The ETIAS CS shall allow ETIAS CU to request additional information of a risk profile to a ETIAS NU. VIS framework.
Description	The ETIAS Central System (CS) shall allow ETIAS Central Unit (CU) to request analysis of a risk profile from the ETIAS NU. The analysis and any review of it shall include at least: (a) a description of the security risk or threat identified, including the frequency, trends and impacts of past occurrences of incidents, or the risks associated with abnormal rates of overstaying or refusals of entry identified, including the frequency, trends and impacts of past occurrences of incidents, specifying a unique reference number for each risk or threat; (b) a list of known facts, evidence and intelligence related to the security risk or threat identified or evidence concerning: (i) an abnormal rate or absolute number of overstayers for a specific group of visa holders present in that Member State or in a specific area of that Member State compared to other specific groups of visa holders present in that Member State or in a specific area of that Member State; or (ii) an abnormal rate or absolute number of refusals of entry issued by that Member State for a specific group of visa holders compared to the number of refusals of entry issued by that Member State for other specific groups of visa holders; or (c) the sets of characteristics of specific groups of visa holders[94] potentially posing or being particularly exposed to the security risk or threat identified or those associated with overstaying or refusal of entry potentially posing or being particularly exposed to the risk identified, considering the following data: (i) age range, sex, nationality; (ii) country and city of residence; (iii) level of education (primary, secondary, higher or none); (iv) current occupation (job group). Where one or more of the elements referred to in point (a) are not available, a justification shall be provided by Member States to ETIAS

[94] In compliance with the principles established in the EU fundamental rights, the sets of characteristics of specific group of travellers shall not lead to direct or indirect discrimination

RP Specific ID	REQ-04.01.05 (SHR)
	CU. Whenever a request is saved, the system shall update the risk indicator and send a notification to the ETIAS NU.
Status	Postponed
Source	01: Annex 8.2 ETIAS Business Use Case Model 02: ETIAS Risk Screening BRD 03: X(2021)-20211217 Definition of risks DA
Rationale	01: Sec. 9.5.14 02: Sec. 4.1(3.1.5) 03: Art.4(1)(a)(b)(c), Art.4(2), Art. 5(1)(a)(b)(c), Art.5(2)
Keywords	request additional information; risk profile; notification
Priority	Should Have
Target / Sub target	ETIAS/Screening Rule
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.4.2.6 REQ-04.01.06 (SHR): The ETIAS CU, in consultation with the ETIAS NU of relevant MSs and the ECDC, shall conduct different analysis of rates, security risk and threats. ETIAS framework.

RP Specific ID	REQ-04.01.06 (SHR)
Requirement ID	AIC-REQ-00059
Name	The ETIAS CU, in consultation with the ETIAS NU of relevant MSs and the ECDC, shall conduct different analysis of rates, security risk and

RP Specific ID	REQ-04.01.06 (SHR)
	threats. ETIAS framework.
Description	The ETIAS CU, in consultation with the ETIAS NU of relevant MSs and the ECDC, shall conduct: (a) analysis of the reasons behind specific abnormal rates or absolute numbers identified for specific groups of persons associated with overstaying, refusal of entry or refusal of TA; (b) periodical analysis on specific security risks or threats; (c) analysis of abnormal rates of overstaying and refusals of entry; (d) analysis of high epidemic risks. For the purposes of the analysis, the software shall enable ETIAS NUs and the ETIAS CU to exchange information. The information shall be provided in a specific format, considering the provisions related to the format of personal data and the ones related to the predetermined list of job groups.
Status	Proposed
Source	01: C(2021)4981-20211123 ETIAS Risks DA
Rationale	01: Art. 3(3), Art. 4, Art. 5, Art. 6(2), Art. 7
Keywords	risk; indicators; screening rules; information; exchange
Priority	Should Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.4.2.7 REQ-04.01.07 (SHR): The ETIAS CU, in consultation with the ETIAS NU of relevant MSs and the ECDC, shall carry out different analysis of rates, security risk and threats. VIS framework.

RP Specific ID	REQ-04.01.07 (SHR)
Requirement ID	AIC-REQ-00064
Name	The ETIAS CU, in consultation with the ETIAS NU of relevant MSs and the ECDC, shall carry out different analysis of rates, security risk and threats. VIS framework.
Description	The ETIAS CU, in consultation with the central visa authority of relevant MSs and the ECDC, shall carry out: (a) analysis of the reasons behind specific abnormal rates or absolute numbers identified for specific groups of visa holders associated with overstaying, refusal of entry or refusal of TA; (b) periodical analysis on specific security risks or threats; (c) analysis of abnormal rates of overstaying and refusals of entry; (d) analysis of high epidemic risks. For the purposes of the analysis, the software shall enable visa authority of relevant MSs and the ETIAS CU to exchange information. The information shall be provided in a specific format, considering the provisions related to the format of personal data and the ones related to the predetermined list of job groups.
Status	Postponed
Source	01: X(2021)-20211217 Definition of risks DA
Rationale	01: Art. 3(3), Art. 4, Art. 5, Art. 6(2), Art. 7
Keywords	risk; indicators; screening rules; information; exchange
Priority	Should Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager

RP Specific ID	REQ-04.01.07 (SHR)
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.5 RP.05 Risk profile validity assessment.

9.5.1 Business Requirements (RP.05)

9.5.1.1 REQ-05.01 (BR): The CRRS shall provide statistics for the purpose in further defining risks related to security or illegal immigration or a high epidemic risk in relation to the ETIAS information system.

RP Specific ID	REQ-05.01 (BR)
Requirement ID	AIC-REQ-00012
Name	The CRRS shall provide statistics for the purpose in further defining risks related to security or illegal immigration or a high epidemic risk in relation to the ETIAS information system.
Description	The CRRS shall provide the following statistics for the purpose of further defining risks related to security or illegal immigration or a high epidemic risk in relation to the ETIAS information system: a) statistics generated by the EES indicating abnormal rates of overstaying and refusals of entry for a specific group of travellers; b) statistics generated by ETIAS indicating abnormal rates of refusals of travel authorisations due to a security, illegal immigration or high epidemic risk associated with a specific group of travellers; c) statistics generated by ETIAS and the EES indicating correlations between information collected through the application form and overstaying by travellers or refusals of entry.
Status	Proposed
Source	01: 02018R1240-20210803 ETIAS REG
Rationale	01: Art. 33(2)

RP Specific ID	REQ-05.01 (BR)
Keywords	ETIAS; exercise; delegation; risks; security; illegal; immigration; epidemic; FRONTEX
Priority	Must have
Target / Sub target	
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.1.2 REQ-05.02 (BR): The ETIAS IS shall provide statistics to further define risks related to security or illegal immigration or a high epidemic risk.

RP Specific ID	REQ-05.02 (BR)
Requirement ID	AIC-REQ-00007
Name	The ETIAS IS shall provide statistics to further define risks related to security or illegal immigration or a high epidemic risk.
Description	In order for the ETIAS CU to further define the risks related to security or illegal immigration or a high epidemic risk, the ETIAS CS shall enable to generate and/or access statistics indicating: (a) overstaying and refusals of entry for a specific group of travellers in EES; (b) refusals of TAs due to a security, illegal immigration or high epidemic risk associated with a specific group of travellers; (c) correlations between information collected in ETIAS IS through the application form and alerts in EES related to overstaying by travellers or refusals of entry.
Status	Proposed

RP Specific ID	REQ-05.02 (BR)
Source	01: 02018R1240-20210803 ETIAS REG
Rationale	01: Art. 33(2)(a), 33(2)(b)
Keywords	data; refusal; entry; overstay; risk; access; EES; security; illegal immigration; epidemic
Priority	Must have
Target / Sub target	
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.1.3 REQ-05.03 (BR): The CRRS shall provide statistics for the purpose of defining the risks related to security or illegal immigration or a high epidemic risk, in the context of the VIS.

RP Specific ID	REQ-05.03 (BR)
Requirement ID	AIC-REQ-00011
Name	The CRRS shall provide statistics for the purpose of defining the risks related to security or illegal immigration or a high epidemic risk, in the context of the VIS.
Description	The CRRS shall provide the following statistics for the purpose defining the risks related to security or illegal immigration or a high epidemic risks in the context of the VIS: (a) statistics generated by the EES indicating abnormal rates of overstaying and refusals of entry for a specific group of visa holders; (b) statistics generated by the VIS in indicating abnormal rates of refusals of

RP Specific ID	REQ-05.03 (BR)
	visa applications due to a security, illegal immigration or high epidemic risk associated with a specific group of visa holders; (c) statistics generated by the VIS and the EES indicating correlations between information collected through the application form and overstaying by visa holders or refusals of entry.
Status	Proposed
Source	01: 32021R1134-20210713 VIS REG
Rationale	01: Art. 9j(2)
Keywords	risk; security; illegal immigration; epidemic; statistics; EES; VIS; visa holder; application refusal; overstaying
Priority	Must have
Target / Sub target	AI in CRRS\Intake
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.2 Stakeholder Requirements (RP.05)

9.5.2.1 REQ-05.00.01 (SHR): Every 6 months or whenever necessary, the ETIAS CU shall assess the continued validity of the risks on "the VIS list".

RP Specific ID	REQ-05.00.01 (SHR)
Requirement ID	AIC-REQ-00068
Name	Every 6 months or whenever necessary, the ETIAS CU shall assess the continued validity of the risks on "the VIS list".
Description	Every 6 months or whenever necessary, the ETIAS CU shall assess the continued validity of the risks on "the VIS list". It shall review and remove from the list risks considered no longer valid.
Status	Postponed
Source	01: X(2021)-20211217 Specification of risks according to Art. 9j3 IA
Rationale	01: Art. 4(1)
Keywords	assess; risk profile; confirm; periodical;
Priority	Should Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.5.2.2 REQ-05.00.02 (SHR): Every 6 months or whenever necessary, the ETIAS CU shall assess the continued validity of the risks on "the list".

RP Specific ID	REQ-05.00.02 (SHR)
Requirement ID	AIC-REQ-00056
Name	Every 6 months or whenever necessary, the ETIAS CU shall assess the continued validity of the risks on "the list".
Description	Every 6 months or whenever necessary, the ETIAS CU shall assess the continued validity of the risks on "the list". It shall review and remove from the list risks considered no longer valid.
Status	Postponed
Source	01: X(2022)-20220304 Specification of risks IA
Rationale	01: Art. 4(1)
Keywords	assess; risk profile; confirm; periodical;
Priority	Should Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.5.2.3 REQ-05.01.01 (SHR): The CRRS shall provide statistics to the ETIAS CU allowing to define risk indicators in consultation with the ETIAS Screening Board. VIS framework.

RP Specific ID	REQ-05.01.01 (SHR)
Requirement ID	AIC-REQ-00066
Name	The CRRS shall provide statistics to the ETIAS CU allowing to define risk indicators in consultation with the ETIAS Screening Board. VIS framework.
Description	To further defining the risks related to security or illegal immigration or a high epidemic risk on the basis, the CRRS shall enable the ETIAS Central Unit to obtain reports related to abnormal rates of overstaying, refusal of entry, refusal of visa applications, and correlations between EES and ETIAS data. The CRRS shall further allow the ETIAS Central Unit to obtain statistical data, statistical reports and customisable reports enabling analysis of those data and reports to further define the risks related to overstaying, refusal of entry and refusal of visa applications, by identifying: a) the total number of cases associated with overstaying, refusal of entry or refusal of visa applications, respectively, using historical data; b) specific groups of visa holders with overstaying, refusal of entry or refusal of visa holders, respectively, based on data derived from abnormal rates or absolute numbers of overstayers, refusals of entry or refusals of visa holders, respectively, compared to other specific groups; c) the sets of characteristics[95] corresponding to the specific groups of visa holders identified under point (b), d) the sets of characteristics corresponding to correlations between the specific groups of visa holders identified under point (b) and information collected through the application file. Definitions: - statistical data: means all the data contained in the central repository for reporting and statistics, which is anonymised and used solely for the purpose of producing statistical reports according to the regulations on the EU information systems. - statistical reports: means an organised collection of statistical data, produced by the central repository for reporting and statistics in an automated manner according to a set of pre-established rules and stored in the central repository. - customisable reports: are statistical reports which are extracted based

[95] In compliance with the principles established in the EU fundamental rights, the sets of characteristics of specific group of travellers shall not lead to direct or indirect discrimination

RP Specific ID	REQ-05.01.01 (SHR)
	on the statistical data contained in the central repository for reporting and statistics according to specific rules defined by a user for an ad-hoc purpose, and which may be stored in the central repository.
Status	Postponed
Source	01: (2021)-20211217 Definition of risks DA 01: 32021R2222-20210930 Operation of CRRS supplementing 2019/818 DA 03: CDR CRRS Reg 2019 817
Rationale	01: Art. 3(2) 02: Art. 1(1), 1(2), 1(3) 03: Art. 1(1), 1(2), 1(3)
Keywords	EES; ETIAS; statistics; risks; overstaying; refusal; entry; travel authorisation; FRONTEX
Priority	Must Have
Target / Sub target	CRRS/Uptake
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.2.4 REQ-05.01.02 (SHR): The ETIAS IS shall enable access to the CRRS to obtain data relating to overstaying and refusals of entries. VIS framework.

RP Specific ID	REQ-05.01.02 (SHR)
Requirement ID	AIC-REQ-00065
Name	The ETIAS IS shall enable access to the CRRS to obtain data relating to overstaying and refusals of entries. VIS framework.
Description	The ETIAS IS shall grant access to the CRRS to obtain the following data, without allowing for individual identification: (a) data generated by the ETIAS IS indicating abnormal rates of refusals of visa holders due to a security, illegal immigration or high epidemic risk associated with a specific group of visa holders; (b) data generated by the ETIAS IS and the EES indicating correlations between information collected through the application form and overstaying by visa holders or refusals of entry.
Status	Postponed
Source	01: (2021)-20211217 Definition of risks DA 02: 32021R2223-20210930 Operation of CRRS supplementing 2019/817_DA
Rationale	01: Art. 3(2) 02: Art. 2(1), Art. 2(5)(c)
Keywords	anonymised; data; risk; refusal, entry; overstay; correlations
Priority	Should Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.5.2.5 REQ-05.01.03 (SHR): The CRRS shall provide statistical reports to the duly authorised staff of the competent authorities on the records kept in the ETIAS.

RP Specific ID	REQ-05.01.03 (SHR)
Requirement ID	AIC-REQ-00033
Name	The CRRS shall provide statistical reports to the duly authorised staff of the competent authorities on the records kept in the ETIAS.
Description	The CRRS shall provide the following business reports, technical reports and the following statistics to the duly authorised staff of the competent authorities of Member States, the Commission, the ETIAS CU and eu-LISA on records kept in the ETIAS: a) Daily statistics on number and nationality of applicants whose travel authorisation was issued or refused, including the grounds for refusal and of third-country nationals whose travel authorisation was annulled or revoked; b) Customisable reports and statistics to improve the assessment of security, illegal immigration and high epidemic risks, to enhance the efficiency of border checks, to help the ETIAS Central Unit and the ETIAS National Units process travel authorisation applications and to support evidence-based Union migration policymaking; c) Statistical data concerning the ETIAS watchlist, specifically statistical data regarding the mandatory review by Europol and Member States of their watchlist alerts that is to take place at least once a year; d) Regular statistics for ensuring the monitoring by eu-LISA of the development and the functioning of the ETIAS Information System, considering objectives relating to planning and costs regarding the development of the system, and considering objectives relating to the technical output, cost-effectiveness, security and quality of service as regards to the functioning of the system.
Status	Proposed
Source	01: 32021R2223-20210930 Operation of CRRS supplementing 2019/817_DA
Rationale	01: Art. 2(5)(b)
Keywords	ETIAS; statistics; applicants; revoked; watchlist; monitoring; FRONTEX
Priority	Should Have
Target / Sub target	CRRS/Uptake CRRS/Platform Management

RP Specific ID	REQ-05.01.03 (SHR)
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.2.6 REQ-05.01.04 (SHR): The CRRS shall provide statistical reports to the duly authorised staff based on statistics collected for the purpose of further defining risks for the ETIAS screening rules.

RP Specific ID	REQ-05.01.04 (SHR)
Requirement ID	AIC-REQ-00034
Name	The CRRS shall provide statistical reports to the duly authorised staff based on statistics collected for the purpose of further defining risks for the ETIAS screening rules.
Description	The CRRS shall provide statistical reports to the duly authorised staff of the competent authorities of Member States, the ETIAS CU, the Commission and eu-LISA on statistics collected for the purpose of further defining risks for the ETIAS screening rules. Those reports will contain the following statistics: a) statistics generated by the EES indicating abnormal rates of overstaying and refusals of entry for a specific group of travellers; b) statistics generated by ETIAS indicating abnormal rates of refusals of travel authorisations due to a security, illegal immigration or high epidemic risk associated with a specific group of travellers; c) statistics generated by the ETIAS and the EES indicating correlations between information collected through the application form and overstaying by travellers or refusals of entry.
Status	Proposed

RP Specific ID	REQ-05.01.04 (SHR)
Source	01: 32021R2223-20210930 Operation of CRRS supplementing 2019/817_DA
Rationale	01: Art. 2(5)(c)
Keywords	enable; authorised; staff; reports; statistics; EES; ETIAS; FRONTEX
Priority	Should Have
Target / Sub target	CRRS/Uptake
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.2.7 REQ-05.01.05 (SHR): The AI technology shall perform comparative analysis of validated risk profiles performance (ex-post).

RP Specific ID	REQ-05.01.05 (SHR)
Requirement ID	AIC-REQ-00043
Name	The AI technology shall perform comparative analysis of validated risk profiles performance (ex-post).
Description	The AI technology shall perform comparative analysis of several validated risk profiles based on their ex-post assessment results. The AI technology shall identify patterns, trends and behaviours. The system shall allow outputs to be exported.
Status	Proposed

RP Specific ID	REQ-05.01.05 (SHR)
Source	01: Annex 8.2 ETIAS Business Use Case Model 02: ETIAS Risk Screening BRD
Rationale	01: Sec. 9.5.13 02: Sec. 8.2
Keywords	ex-post assessment; risk profile; comparison report
Priority	Should Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.5.2.8 REQ-05.01.06 (SHR): The ETIAS CS shall allow ex-post assessments.

RP Specific ID	REQ-05.01.06 (SHR)
Requirement ID	AIC-REQ-00058
Name	The ETIAS CS shall allow ex-post assessments.
Description	The ETIAS CS shall allow authorised users to perform on demand ex-post assessments of validated risk profiles in an anonymous manner. The assessments shall provide the daily matching rate percentage and plot the evolution of the matching rate. The system shall allow to introduce and save comments to the assessments. The system shall allow ex-post assessment to be exported. The different analyses shall not contain any personal data.
Status	Proposed

RP Specific ID	REQ-05.01.06 (SHR)
Source	01: Annex 8.2 ETIAS Business Use Case Model 02: ETIAS Risk Screening BRD 03: C(2021)4981-20211123 ETIAS Risks DA 04: X(2021)-20211217 Definition of risks DA
Rationale	01: Sec. 9.5.11 02: Sec. 8.1 03: Art. 8 04: Art. 8
Keywords	ex-post assessment; risk profile; data protection
Priority	Should Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.5.2.9 REQ-05.01.07 (SHR): The CRRS shall provide statistics to the ETIAS CU allowing to define risk indicators in consultation with the ETIAS Screening Board. ETIAS framework.

RP Specific ID	REQ-05.01.07 (SHR)
Requirement ID	AIC-REQ-00036
Name	The CRRS shall provide statistics to the ETIAS CU allowing to define risk indicators in consultation with the ETIAS Screening Board. ETIAS framework.
Description	To further defining the risks related to security or illegal immigration or a high epidemic risk on the basis, the CRRS shall enable the ETIAS Central Unit to obtain reports related to abnormal rates of overstaying, refusal of entry, refusal of travel authorisation and correlations between

RP Specific ID	REQ-05.01.07 (SHR)
	EES and ETIAS data. The CRRS shall further allow the ETIAS Central Unit to obtain statistical data, statistical reports and customisable reports enabling analysis of those data and reports to further define the risks related to overstaying, refusal of entry and refusal of travel authorisation, by identifying: a) the total number of cases associated with overstaying, refusal of entry or refusal of travel authorisation, respectively, using historical data; b) specific groups of travellers associated with overstaying, refusal of entry or refusal of travel authorisation, respectively, based on data derived from abnormal rates or absolute numbers of overstayers, refusals of entry or refusals of travel authorisation, respectively, compared to other specific groups; c) the sets of characteristics[96] corresponding to the specific groups of travellers identified under point (b), d) the sets of characteristics corresponding to correlations between the specific groups of travellers identified under point (b) and information collected through the application file. Definitions: - statistical data: means all the data contained in the central repository for reporting and statistics, which is anonymised and used solely for the purpose of producing statistical reports according to the regulations on the EU information systems. - statistical reports: means an organised collection of statistical data, produced by the central repository for reporting and statistics in an automated manner according to a set of pre-established rules and stored in the central repository. - customisable reports: are statistical reports which are extracted based on the statistical data contained in the central repository for reporting and statistics according to specific rules defined by a user for an ad-hoc purpose, and which may be stored in the central repository.
Status	Proposed
Source	01: C(2021)4981-20211123 ETIAS Risks DA 01: 32021R2222-20210930 Operation of CRRS supplementing 2019/818 DA 03: CDR CRRS Reg 2019 817

[96] In compliance with the principles established in the EU fundamental rights, the sets of characteristics of specific group of travellers shall not lead to direct or indirect discrimination

RP Specific ID	REQ-05.01.07 (SHR)
Rationale	01: Art. 3(2) 02: Art. 1(1), 1(2), 1(3) 03: Art. 1(1), 1(2), 1(3)
Keywords	EES; ETIAS; statistics; risks; overstaying; refusal; entry; travel authorisation; FRONTEX
Priority	Must Have
Target / Sub target	CRRS/Uptake
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.2.10 REQ-05.01.08 (SHR): The ETIAS IS shall enable access to the CRRS to obtain data relating to overstaying and refusals of entries. ETIAS framework.

RP Specific ID	REQ-05.01.08 (SHR)
Requirement ID	AIC-REQ-00060
Name	The ETIAS IS shall enable access to the CRRS to obtain data relating to overstaying and refusals of entries. ETIAS framework.
Description	The ETIAS IS shall grant access to the CRRS to obtain the following data, without allowing for individual identification: (a) data generated by the ETIAS IS indicating abnormal rates of refusals of TAs due to a security, illegal immigration or high epidemic risk associated with a specific group of travellers; (b) data generated by the ETIAS IS and the EES indicating correlations between information collected through the application form and overstaying by travellers or refusals of entry.

RP Specific ID	REQ-05.01.08 (SHR)
Status	Proposed
Source	01: C(2021)4981-20211123 ETIAS Risks DA 02: CDR CRRS Reg 2019 817
Rationale	01: Art. 3(2) 02: Art. 2(1), Art. 2(5)(c)
Keywords	anonymised; data; risk; refusal, entry; overstay
Priority	Should Have
Target / Sub target	ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.5.2.11 REQ-05.01.09 (SHR): The CRRS shall provide statistics to the ETIAS Central Unit allowing to define risk indicators in consultation with the VIS Screening Board.

RP Specific ID	REQ-05.01.09 (SHR)
Requirement ID	AIC-REQ-00067
Name	The CRRS shall provide statistics to the ETIAS Central Unit allowing to define risk indicators in consultation with the VIS Screening Board.
Description	The CRRS shall allow the ETIAS Central Unit to obtain statistical data, statistical reports and customisable reports enabling analysis of data and reports to further define the risks related to overstaying, refusal of entry and refusal of visa applications. The data, statistical reports, customisable reports and functionalities supporting the creation of such reports shall be such that the ETIAS

RP Specific ID	REQ-05.01.09 (SHR)
	Central Unit is able to assess what should constitute a normal and abnormal rate by applying the business rules for this process. Those rules are the following; a) the specific time-period considered when assessing whether a rate or absolute number is normal or abnormal shall be three months. This period shall be compared with a similar time period from the previous year or years or, where necessary due to the nature of the event, to any previous year or years; b) data allowing to determine specific thresholds for considering a rate or absolute number for the specific time-period in question normal; c) where either the rate or the absolute number is above or below the specific thresholds determined in accordance with point (b), the rate or absolute number shall be considered abnormal, representing a heightened risk, unless the rate or absolute number in question can be explained by other factors, including: 1) periodical variations in line with historical observations, including seasonality; 2) modification of border management processes or systems, including due to new legislation or new technology; 3) stricter border control checks, including due to joint or national operations; 4) singular events, including international sport events or national elections. d) For the purpose of applying point (c), the CRRS shall allow to specify acceptable error margins. The customisation of reports shall allow for the ETIAS Central Unit to use other rules to assess normal and abnormal rates than the rules described in point a)-d), if it is necessary for reasons related to the risk in question.
Status	Postponed
Source	01: X(2021)-20211217 Specification of risks according to Art. 9j3 IA
Rationale	01: Art. 3(2)
Keywords	ETIAS; risks; screening; rules; FRONTEX
Priority	Must Have
Target / Sub target	CRRS/Uptake

RP Specific ID	REQ-05.01.09 (SHR)
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.2.12 REQ-05.01.10 (SHR): The CRRS shall provide statistics to the ETIAS Central Unit allowing to define risk indicators in consultation with the ETIAS Screening Board.

RP Specific ID	REQ-05.01.10 (SHR)
Requirement ID	AIC-REQ-00037
Name	The CRRS shall provide statistics to the ETIAS Central Unit allowing to define risk indicators in consultation with the ETIAS Screening Board.
Description	The CRRS shall allow the ETIAS Central Unit to obtain statistical data, statistical reports and customisable reports enabling analysis of data and reports to further define the risks related to overstaying, refusal of entry and refusal of travel authorisation. The data, statistical reports, customisable reports and functionalities supporting the creation of such reports shall be such that the ETIAS Central Unit is able to assess what should constitute a normal and abnormal rate by applying the business rules for this process. Those rules are the following; a) the specific time-period considered when assessing whether a rate or absolute number is normal or abnormal shall be three months. This period shall be compared with a similar time period from the previous year or years or, where necessary due to the nature of the event, to any previous year or years; b) data allowing to determine specific thresholds for considering a rate or absolute number for the specific time-period in question normal; c) where either the rate or the absolute number is above or below the specific thresholds determined in accordance with point (b), the rate or absolute number shall be considered abnormal, representing a

RP Specific ID	REQ-05.01.10 (SHR)
	heightened risk, unless the rate or absolute number in question can be explained by other factors, including: 1) periodical variations in line with historical observations, including seasonality; 2) modification of border management processes or systems, including due to new legislation or new technology; 3) stricter border control checks, including due to joint or national operations; 4) singular events, including international sport events or national elections. d) For the purpose of applying point (c), the CRRS shall allow to specify acceptable error margins. The customisation of reports shall allow for the ETIAS Central Unit to use other rules to assess normal and abnormal rates than the rules described in point a)-d), if it is necessary for reasons related to the risk in question.
Status	Postponed
Source	01: X(2022)-20220304 Specification of risks IA
Rationale	01: Art. 3(2)
Keywords	ETIAS; risks; screening; rules; FRONTEX
Priority	Must Have
Target / Sub target	CRRS/Uptake
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.2.13 REQ-05.02.01 (SHR): The CRRS shall ensure that all data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved.

RP Specific ID	REQ-05.02.01 (SHR)
Requirement ID	AIC-REQ-00088
Name	The CRRS shall ensure that all data will be anonymized and no personal/ sensitive information for the individuals is able to be retrieved.
Description	The CRRS shall provide analytical capabilities to allow to perform extraction of attributes from all CBS in an anonymous manner, ensuring no personal/sensitive information can be retrieved.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	attributes; extraction; anonymization
Priority	Must have
Target / Sub target	CRRS/AI
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.2.14 REQ-05.02.02 (SHR): The CRRS shall allow analytical capabilities on the anomaly detection levels for the risk profiles (i.e., "valid", "non-valid", commonly agreed with the stakeholders).

RP Specific ID	REQ-05.02.02 (SHR)
Requirement ID	AIC-REQ-00087
Name	The CRRS shall allow analytical capabilities on the anomaly detection levels for the risk profiles (i.e., "valid", "non-valid", commonly agreed with the stakeholders).
Description	The CRRS shall provide analytical capabilities to allow to render and merge data from all CBS (EES, ETIAS, VIS, EURODAC, ECRIS TCN, VIS and SIS II) and set up different labels such as "valid" or "non-valid" (or as agreed among eu-Lisa's stakeholders) for a particular risk profile.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	labels; risk; profile
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.5.2.15 REQ-05.02.03 (SHR): The CRRS shall allow a precise definition of the analysis time-interval (e.g., 6-months or less) for the ex-post assessment analysis.

RP Specific ID	REQ-05.02.03 (SHR)
Requirement ID	AIC-REQ-00086
Name	The CRRS shall allow a precise definition of the analysis time-interval (e.g., 6-months or less) for the ex-post assessment analysis.
Description	The CRRS shall provide analytical capabilities to allow to perform time-interval analysis of 6 month or as agreed among eu-LISA stakeholders for the purpose of producing ex-post assessments.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	time; interval; ex-post assessment
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.6 RP.06 Virtual assistance to provide guidance on certain activities.

9.6.1 Business Requirements (RP.06)

9.6.1.1 REQ-06.01 (BR): The CRRS shall allow the competent authorities of Member States, the Commission, eu-LISA and the ETIAS Central Unit to obtain customisable reports and statistics.

RP Specific ID	REQ-06.01 (BR)
Requirement ID	AIC-REQ-00013
Name	The CRRS shall allow the competent authorities of Member States, the Commission, eu-LISA and the ETIAS Central Unit to obtain customisable reports and statistics.
Description	The CRRS shall allow the competent authorities of Member States, the Commission, eu-LISA and the ETIAS Central Unit to obtain customisable reports and statistics, to support the implementation of the ETIAS screening rules, to improve the assessment of the security, illegal immigration and high epidemic risks, to enhance the efficiency of border checks and to help the ETIAS Central Unit and the ETIAS National Units process travel authorisation applications.
Status	Proposed
Source	01: 02018R1240-20210803 ETIAS REG
Rationale	01: Art. 84(2)
Keywords	authorities; customisable; reports; ETIAS; FRONTEX
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled, and it becomes 'true'.

RP Specific ID	REQ-06.01 (BR)
Sector	(BRMS) Business Relations Management Sector

9.6.2 Stakeholder Requirements (RP.06)

9.6.2.1 REQ-06.01.01 (SHR): The CRRS shall provide a catalogue of reports.

RP Specific ID	REQ-06.01.01 (SHR)
Requirement ID	AIC-REQ-00031
Name	The CRRS shall provide a catalogue of reports.
Description	The CRRS shall provide a catalogue of reports. It shall be possible to request new reports or changes to existing ones following and in accordance with eu-LISA change management policy.
Status	Proposed
Source	01: 32021R2223-20210930 Operation of CRRS supplementing 2019/817_DA 01: 32021R2222-20210930 Operation of CRRS supplementing 2019/818 DA
Rationale	01: Art. 2(8) 02: Art. 2(8)
Keywords	catalogue; reports; request; new; changes
Priority	Should Have
Target / Sub target	CRRS/Uptake CRRS/Platform Management
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised

RP Specific ID	REQ-06.01.01 (SHR)
	and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.6.2.2 REQ-06.01.02 (SHR): The CRRS shall include a precise definition on the text data format used as input for the chatbot.

RP Specific ID	REQ-06.01.02 (SHR)
Requirement ID	AIC-REQ-00089
Name	The CRRS shall include a precise definition on the text data format used as input for the chatbot.
Description	The CRRS shall provide analytical capabilities to allow to analyse message sent to ETIAS chatbot through Natural Language Understanding to provide support and to improve critical business processes when and as required.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	text; data; chatbot; guidance; processes
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder

RP Specific ID	REQ-06.01.02 (SHR)
	requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.6.2.3 REQ-06.01.03 (SHR): The CRRS shall provide a virtual assistant that contemplates all ways a conversation can go.

RP Specific ID	REQ-06.01.03 (SHR)
Requirement ID	AIC-REQ-00094
Name	The CRRS shall provide a virtual assistant that contemplates all ways a conversation can go.
Description	The CRRS shall provide analytical capabilities to allow a virtual assistant to contemplate all ways a conversation can take place.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	language; variants; virtual assistant; conversation
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.

RP Specific ID	REQ-06.01.03 (SHR)
Sector	(BRMS) Business Relations Management Sector

9.6.2.4 REQ-06.01.04 (SHR): The CRRS shall include different variants of requests that are similar, and other variants that can be treated differently.

RP Specific ID	REQ-06.01.04 (SHR)
Requirement ID	AIC-REQ-00093
Name	The CRRS shall include different variants of requests that are similar, and other variants that can be treated differently.
Description	The CRRS shall allow analytical capabilities to separate variants of requests from similar to different so that it can perform the purpose that for which it was built.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	language; variants
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.6.2.5 REQ-06.01.05 (SHR): The CRRS shall include a precise definition of the required answers/ actions that need to be extracted through the virtual assistant/ chatbot tool.

RP Specific ID	REQ-06.01.05 (SHR)
Requirement ID	AIC-REQ-00092
Name	The CRRS shall include a precise definition of the required answers/ actions that need to be extracted through the virtual assistant/ chatbot tool.
Description	The CRRS shall allow analytical capabilities to provide a precise definition of the answers to achieve the purpose that for which it was built.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	language; definition; answers
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.6.2.6 REQ-06.01.06 (SHR): The CRRS shall include a precise definition of the questions that need to be learned via the AI technologies.

RP Specific ID	REQ-06.01.06 (SHR)
Requirement ID	AIC-REQ-00091
Name	The CRRS shall include a precise definition of the questions that need to be learned via the AI technologies.
Description	The CRRS shall allow analytical capabilities to provide a precise definition of the questions to achieve the purpose that for which it was built.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	language; definition; questions
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.6.2.7 REQ-06.01.07 (SHR): The CRRS shall include a precise definition on the coding language and required packages/libraries.

RP Specific ID	REQ-06.01.07 (SHR)
Requirement ID	AIC-REQ-00090
Name	The CRRS shall include a precise definition on the coding language and required packages/libraries.
Description	The CRRS shall provide analytical capabilities to allow to apply the required coding and packages or libraries to achieve the purpose that for which it was built.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	language; coding; libraries
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.7 RP.07 Verification process when a hit takes place.

9.7.1 Business Requirements (RP.07)

9.7.1.1 REQ-07.01 (BR): The ETIAS CS shall compare personal data to specific risk indicators to identify hit(s).

RP Specific ID	REQ-07.01 (BR)
Requirement ID	AIC-REQ-00021
Name	The ETIAS CS shall compare personal data to specific risk indicators to identify hit(s).
Description	The ETIAS CS shall compare relevant personal data provided by the applicant in the application form* to the specific risk indicators, established by the ETIAS CU. * (a) surname (family name), first name(s) (given name(s)), surname at birth; date of birth, place of birth, sex, current nationality; (b) country of birth, first name(s) of the parents of the applicant; (c) other nationalities, if any; (d) the applicants home address or, if not available, his or her city and country of residence; (e) education (primary, secondary, higher or none); (f) current occupation (job group).
Status	Accepted
Source	01: 02018R1240-20210803 ETIAS REG 02: 02019R0817-20210803 Interoperability Visa Borders REG v1
Rationale	01: Art. 20(5)
Keywords	ETIAS CS; automated processing; compare; personal data; application form; verify; hit; risk indicators
Priority	Must have
Target / Sub target	ETIAS/Automated Processing
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)

RP Specific ID	REQ-07.01 (BR)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower-level related requirements have been realised and the description of the business requirement has been fulfilled and becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.7.2 Stakeholder Requirements (RP.07)

9.7.2.1 REQ-07.01.01 (SHR): The CRRS shall ensure enough historical data to support human decision-making during verification process on hits.

RP Specific ID	REQ-07.01.01 (SHR)
Requirement ID	AIC-REQ-00097
Name	The CRRS shall ensure enough historical data to support human decision-making during verification process on hits.
Description	The CRRS will execute its analytical capabilities to support human decision-making during verification process on hits when enough data is available from all CBS (EES, ETIAS, VIS, EURODAC, ECRIS TCN, VIS and SIS II).
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	information; CBS; processes; hit
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)

RP Specific ID	REQ-07.01.01 (SHR)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.7.2.2 REQ-07.01.02 (SHR): The CRRS will include anomaly detection levels corresponding to the ETIAS risk profiles and as agreed among eu-LISA's stakeholders.

RP Specific ID	REQ-07.01.02 (SHR)
Requirement ID	AIC-REQ-00096
Name	The CRRS will include anomaly detection levels corresponding to the ETIAS risk profiles and as agreed among eu-LISA's stakeholders.
Description	The CRRS shall provide analytical capabilities to allow to generate Machine Learning-based models on the characteristics of the risk profiles likely to produce hits and no hits, and label them as "hit-candidates"/"no hit candidates" respectively.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	ETIAS CU; anomaly; detection; attributes; hit; no hit
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)

RP Specific ID	REQ-07.01.02 (SHR)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.7.2.3 REQ-07.01.03 (SHR): The CRRS shall include a precise definition of the historical analysis time-window for hit classification.

RP Specific ID	REQ-07.01.03 (SHR)
Requirement ID	AIC-REQ-00095
Name	The CRRS shall include a precise definition of the historical analysis time-window for hit classification.
Description	The CRRS shall provide analytical capabilities to allow to implement Model Training on historical data of corresponding risk profiles to help identify and learn "hit" and "no hit" values.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Interim_Report_00_00_01
Rationale	01: N/A
Keywords	attributes; hit; no hit; risk profile; model training
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager

RP Specific ID	REQ-07.01.03 (SHR)
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8 RP.50 Reusable requirements.

9.8.1 Business Requirements (RP.50)

9.8.1.1 REQ-50.01 (BR): The CRRS shall not receive or process any personal data that could allow discrimination against persons on any grounds or violating human dignity, integrity and fundamental rights.

RP Specific ID	REQ-50.01 (BR)
Requirement ID	AIC-REQ-00019
Name	The CRRS shall not receive or process any personal data that could allow discrimination against persons on any grounds or violating human dignity, integrity and fundamental rights.
Description	The CRRS processing of personal data by any user shall not result in discrimination against third-country nationals on the grounds of sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation. It shall fully respect human dignity and integrity and fundamental rights, including the right to respect for one's private life and to the protection of personal data. Particular attention shall be paid to children, the elderly and persons with a disability. The best interests of the child shall be a primary consideration.
Status	Proposed
Source	01: 02019R0817-20210803 IO Visa Borders REG v1 02: 02019R0818-20210803 IO LEA Immigration REG v1 03: 02006R1987-20201228 SIS 04: 02017R2226-20210803 EES 05: 02018R1240-20210803 ETIAS 06: 32021R1134-20210713 VIS 07: 02019R0816-20210803 ECRIS-TCN
Rationale	01: Art. 5 02: Art. 5 03: Art. 40, 60(15), 61(13) 04: Art. 63(1), 63(2) 05: Art. 14 06: Art. 45a(1), 45a(2) 07: Art. 32(1), 32(2)
Keywords	data management; processing; personal; discrimination; sex; race; colour; ethnic; origin; genetic feature; language; religion; belief; political;

RP Specific ID	REQ-50.01 (BR)
	minority; disability; age; children; SIS; Eurodac; ECRIS-TCN; EES; VIS; ETIAS; CIR; MID; sBMS; ESP
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.1.2 REQ-50.02 (BR): The CRRS shall adhere to the detailed rules laid down in the Commission Delegated Regulation on the operation of the CRRS

RP Specific ID	REQ-50.02 (BR)
Requirement ID	AIC-REQ-00020
Name	The CRRS shall adhere to the detailed rules laid down in the Commission Delegated Regulation on the operation of the CRRS.
Description	The CRRS shall adhere to the detailed rules laid down in the Commission Delegated Regulation on the operation of the CRRS related to the EES, SIS, ETIAS, ECRIS-TCN, VIS, Eurodac and IO components, including the specific safeguards for the processing of personal data and security rules applicable to the repository. The Commission Delegated Regulation includes details on: - The information needed for reporting and statistics, - The data repository and reporting tool, - The extraction of the information, - The anonymisation tool, - The access to the CRRS, - The data processor.

RP Specific ID	REQ-50.02 (BR)
Status	Proposed
Source	01: 02019R0817-20210803 Interoperability Visa Borders REG v1 02: 02019R0818-20210803 Interoperability LEA Immigration REG v1
Rationale	01: Art. 39(5) 02: Art. 39(5)
Keywords	data management; operation; rules; safeguards; processing security; operations; rules; CBSs; EES; SIS; ETIAS; ECRIS-TCN; VIS; Eurodac;
Priority	Must have
Target / Sub target	CRRS/Data Management
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g., SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2 Stakeholder Requirements (RP.50)

9.8.2.1 REQ-50.00.01 (SHR): The AI technology shall consider the existing technological and architectural landscape in eu-LISA.

RP Specific ID	REQ-50.00.01 (SHR)
Requirement ID	AIC-REQ-00072
Name	The AI technology shall consider the existing technological and architectural landscape in eu-LISA.
Description	The AI technology shall consider the existing and future technological and architectural landscape in eu-LISA. The AI technology should be

RP Specific ID	REQ-50.00.01 (SHR)
	aligned with roadmap for the following years.
Status	Proposed
Source	01: AI-CRRS_MoM_Technical Workshop_13052022.docx
Rationale	01: Topic 4
Keywords	AI; technology; COTS; landscape; roadmap
Priority	Must Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.2 REQ-50.00.02 (SHR): The AI technology shall be implemented by means of Open-Source solution(s).

RP Specific ID	REQ-50.00.02 (SHR)
Requirement ID	AIC-REQ-00071
Name	The AI technology shall be implemented by means of Open-Source solution(s).
Description	The AI technology shall not include development or tailor-made solutions, it shall be implemented by means of Open-Source solution(s).

RP Specific ID	REQ-50.00.02 (SHR)
Status	Proposed
Source	01: AI-CRRS_MoM_Technical Workshop_13052022.docx
Rationale	01: Topic 4
Keywords	AI; technology; open source; solution
Priority	Must Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.3 REQ-50.00.03 (SHR): When identifying a potential risk profile, the AI technology shall provide a Description, Indicators, Screening Rule or Matching Threshold.

RP Specific ID	REQ-50.00.03 (SHR)
Requirement ID	AIC-REQ-00040
Name	When identifying a potential risk profile, the AI technology shall provide a Description, Indicators, Screening Rule or Matching Threshold.
Description	The AI technology shall provide the findings following the existing structure including Description, Indicators, Screening Rule or Matching Threshold. The system shall allow the ETIAS CU user to modify the proposed values. It is necessary to confirm all information before saving a risk profile.

RP Specific ID	REQ-50.00.03 (SHR)
Status	Proposed
Source	01: Annex 8.2 ETIAS Business Use Case Model
Rationale	01: Sec. 9.5.4(2)
Keywords	description; indicators; screening rule; matching threshold; profile
Priority	Should Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	Lower level functional and non-functional related requirements have been implemented therefore the stakeholder requirement has been fulfilled.
Sector	(BRMS) Business Relations Management Sector

9.8.2.4 REQ-50.00.04 (SHR): The CRRS shall allow the deployment of a natural language interface (NLI).

RP Specific ID	REQ-50.00.04 (SHR)
Requirement ID	AIC-REQ-00050
Name	The CRRS shall allow the deployment of a natural language interface (NLI).
Description	The CRRS shall allow the use of a chatbot-like solution to allow a natural-language search engine to find targeted answers to user questions (as opposed to a keyword search) regarding the use of the reporting tool. When deployed in the CRRS the Natural-language search shall use natural-language processing to understand the nature of the question and then to search and return a subset of the data that contains the answer to the question. The results shall have higher relevance than

RP Specific ID	REQ-50.00.04 (SHR)
	results from a keyword search engine, due to the question being included.
Status	Proposed
Source	01: Annex 13.1 CRRS Technical specifications Requirements 02: Annex 6.0 CRRS Requirements
Rationale	01: Sec. 6.6.1 02: Sec. 7.3.21
Keywords	reporting; artificial intelligence; natural; language; interface
Priority	Should Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.5 REQ-50.00.05 (SHR): The AI technology shall not be used to predict or forecast the future.

RP Specific ID	REQ-50.00.05 (SHR)
Requirement ID	AIC-REQ-00052
Name	The AI technology shall not be used to predict or forecast the future.
Description	The AI technology shall not allow the processing of past and current information with the goal of predicting or forecasting the future.

RP Specific ID	REQ-50.00.05 (SHR)
Status	Proposed
Source	01: Annex 13.1 CRRS Technical specifications Requirements 02: Annex 6.0 CRRS Requirements
Rationale	01: Sec. 6.6.2 02: Sec 7.3.13
Keywords	analytics; artificial intelligence; forecast; future; trends
Priority	Must Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.6 REQ-50.00.06 (SHR): The AI technology shall enable to identify patterns within the data stored in the CRRS.

RP Specific ID	REQ-50.00.06 (SHR)
Requirement ID	AIC-REQ-00044
Name	The AI technology shall enable to identify patterns within the data stored in the CRRS.
Description	The AI technology in CRRS shall allow to process large amounts of data and recognise patterns in the data that is analysed. The AI technology shall support finding connections and derive meaning from undefined data to learn complex patterns in large amounts of data, enabling the identification of trends.

RP Specific ID	REQ-50.00.06 (SHR)
Status	Proposed
Source	01: Annex 13.1 CRRS Technical specifications Requirements 02: Annex 6.0 CRRS Requirements
Rationale	01: Sec. 6.6.2 02: Sec. 7.3.9
Keywords	analytics; artificial intelligence; capability; point out; patterns
Priority	Should Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.7 REQ-50.00.07 (SHR): The CRRS shall be one of the sources to be utilised by AI technology to analyse current and past information through machine learning capabilities.

RP Specific ID	REQ-50.00.07 (SHR)
Requirement ID	AIC-REQ-00049
Name	The CRRS shall be one of the sources to be utilised by AI technology to analyse current and past information through machine learning capabilities.
Description	The CRRS shall be one of a variety of sources that is to be utilised by AI technology to analyse current and past information with the help of artificial intelligence and machine learning capabilities.

RP Specific ID	REQ-50.00.07 (SHR)
Status	Proposed
Source	01: Annex 13.1 CRRS Technical specifications Requirements 02: Annex 6.0 CRRS Requirements
Rationale	01: Sec. 6.6.2 02: Sec. 7.3.11
Keywords	analytics; artificial intelligence; sources; current; past; information; predict; future; events
Priority	Should Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.8 REQ-50.00.08 (SHR): The AI technology shall allow to analyse and draw inferences from data.

RP Specific ID	REQ-50.00.08 (SHR)
Requirement ID	AIC-REQ-00045
Name	The AI technology shall allow to analyse and draw inferences from data.
Description	The AI technology shall provide statistical analytic tools that are able to analyse and draw inductive inferences from data. The tools shall apply statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.

RP Specific ID	REQ-50.00.08 (SHR)
Status	Proposed
Source	01: Annex 13.1 CRRS Technical specifications Requirements 02: Annex 6.0 CRRS Requirements
Rationale	01: Sec. 6.6.2 02: Sec. 7.3.12
Keywords	analytics; artificial intelligence; statistical; tools; analyse; data; inferences
Priority	Must Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.9 REQ-50.00.09 (SHR): The AI technology shall be made available so the capabilities to learn based on data inputs and/or outputs extend beyond the specific objectives of this project.

RP Specific ID	REQ-50.00.09 (SHR)
Requirement ID	AIC-REQ-00047
Name	The AI technology shall be made available so the capabilities to learn based on data inputs and/or outputs extend beyond the specific objectives of this project.
Description	The CRRS data shall be used to augment Artificial Intelligence and machine learning capabilities to improve the ability to detect patterns and improving capabilities through analysing input data and/or output

RP Specific ID	REQ-50.00.09 (SHR)
	data.
Status	Proposed
Source	01: Annex 13.1 CRRS Technical specifications Requirements 02: Annex 6.0 CRRS Requirements
Rationale	01: Sec. 6.6.2 02: Sec. 7.3.25
Keywords	analytics; artificial intelligence; learn; data; inputs; outputs
Priority	Could have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.10 REQ-50.00.10 (SHR): The AI technology shall enable capabilities to provide automated mechanisms to predict and to discover potential risks.

RP Specific ID	REQ-50.00.10 (SHR)
Requirement ID	AIC-REQ-00048
Name	The AI technology shall enable capabilities to provide automated mechanisms to predict and to discover potential risks.
Description	The AI technology shall enable capabilities to provide automated mechanisms to predict and to discover potential risks behaviour by

RP Specific ID	REQ-50.00.10 (SHR)
	applying learning models on the different data that is stored in the CRRS.
Status	Proposed
Source	01: Annex 13.1 CRRS Technical specifications Requirements 02: Annex 6.0 CRRS Requirements
Rationale	01: Sec. 6.6.2 02: Sec. 7.3.26
Keywords	analytics; artificial intelligence; detect; predict; discover; fraud
Priority	Could have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.11 REQ-50.00.11 (SHR): The CRRS shall allow the use of a Natural Language user Interface (NLI) for creating, selecting and modifying visualisations of data.

RP Specific ID	REQ-50.00.11 (SHR)
Requirement ID	AIC-REQ-00051
Name	The CRRS shall allow the use of a Natural Language user Interface (NLI) for creating, selecting and modifying visualisations of data.
Description	The CRRS shall allow the deployment of a Natural Language user Interface (NLI) that will be able to analyse, understand and generate

RP Specific ID	REQ-50.00.11 (SHR)
	human language that allows users to interact with the system using a normal everyday language. When deployed in the CRRS it will help users to perform tasks in the CRRS such as customising reports or getting information on how to use the web portal. The primary objective is enabling CRRS to interact with humans the way that humans interact with each other.
Status	Proposed
Source	01: Annex 13.1 CRRS Technical specifications Requirements 02: Annex 6.0 CRRS Requirements
Rationale	01: Sec. 6.6.2 02: Sec. 7.3.28
Keywords	analytics; artificial intelligence; natural; language; user; interfaces; interact; help; assistance
Priority	Must Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Contractor)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.12 REQ-50.00.12 (SHR): The AI technology shall include the implementation of machine learning models.

RP Specific ID	REQ-50.00.12 (SHR)
Requirement ID	AIC-REQ-00053
Name	The AI technology shall include the implementation of machine learning models.
Description	The AI technology shall include deployment of tools enabling the detection of patterns in the data via the implementation of automated analytical model building (Machine Learning Models) in, but not limited to, CRRS.
Status	Proposed
Source	01: Annex 13.1 CRRS Technical specifications Requirements 02: Annex 6.0 CRRS Requirements
Rationale	01: Sec. 6.6.2 02: Sec. 7.4.15
Keywords	analytics; artificial intelligence; patterns; machine learning; models
Priority	Must Have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.13 REQ-50.00.13 (SHR): The AI technology shall include smart re-usable tools, which enable the automated monitoring of data for the purpose of detecting anomalies.

RP Specific ID	REQ-50.00.13 (SHR)
Requirement ID	AIC-REQ-00046
Name	The AI technology shall include smart re-usable tools, which enable the automated monitoring of data for the purpose of detecting anomalies.
Description	The AI technology shall make use of smart re-usable tools allowing to define Machine Learning Models with the aim of monitoring data to detect and report/alert unusual activity or anomalies over time, including but not limited to particular seasons. When deployed in the CRRS, the tools shall identify patterns in the data and learn higher order connections between information to provide insight into whether a given activity (piece of data) fits an existing pattern or if the activity (piece of data) is an outlier and does not fit.
Status	Proposed
Source	01: Annex 13.1 CRRS Technical specifications Requirements 02: Annex 6.0 CRRS Requirements
Rationale	01: Sec. 6.6.2 02: Sec. 7.4.15
Keywords	analytics; artificial intelligence; monitoring; activity; machine learning; models
Priority	Should Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.14 REQ-50.00.14 (SHR): The CRRS shall allow AI technologies to be integrated with the standard eu-LISA current and existing technologies landscape.

RP Specific ID	REQ-50.00.14 (SHR)
Requirement ID	AIC-REQ-00100
Name	The AI technologies in CRRS, ETIAS and revised VIS shall be integrated with standard eu-LISA current & existing technologies.
Description	The AI technologies in CRRS, ETIAS and revised VIS from the architectural, process, storage and infrastructure point of view shall be integrated with the standard eu-LISA current & existing technologies landscape.
Status	Proposed
Source	01: AI-CRRS-E1-SC05-002_-_AI_in_CRRS_Final_Report_00_00_01-DIS_APO
Rationale	01: Sctn. 5.4 (Pg 185 Ln 32) Sctn. 7 (Pg 212 Ln 1)
Keywords	AI; technology; CRRS; ETIAS; revised VIS; architectural; landscape; integration; current; technologies
Priority	Must Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.15 REQ-50.01.01 (SHR): The AI technology shall comply with the regulatory framework for AI.

RP Specific ID	REQ-50.01.01 (SHR)
Requirement ID	AIC-REQ-00070
Name	The AI technology shall comply with the regulatory framework for AI.
Description	The AI technology shall comply with ICDPPC "Declaration on Ethics and Data Protection in AI" and any other regulatory framework. NOTE: AI technology shall be implemented in line with the EU Charter of Fundamental Rights.[97] *ICDPPC: International Conference of Data Protection and Privacy Commissioners (now renamed Global Privacy Assembly)
Status	Proposed
Source	01: EDPS Opinion on the European Commission's White Paper on Artificial Intelligence 02: Declaration on ethics and data protection in AI
Rationale	01: Sec. 3(20) 02: N/A
Keywords	ethics; data protection; human rights; regulatory; framework
Priority	Must Have
Target / Sub target	CRRS/ETIAS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager

[97] https //commission europa eu/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights_en

RP Specific ID	REQ-50.01.01 (SHR)
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this business requirement is fully fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

9.8.2.16 REQ-50.01.02 (SHR): The CRRS shall comply with Accountability Principles for Artificial Intelligence

RP Specific ID	REQ-50.01.02 (SHR)
Requirement ID	AIC-REQ-00098
Name	The CRRS shall comply with Accountability Principles for Artificial Intelligence.
Description	The CRRS shall comply with "Accountability Principles for Artificial Intelligence (AP4AI) in the internal security domain", in particular the 12 principles: 1. Legality 2. Universality 3. Pluralism 4. Transparency 5. Independence 6. Commitment to robust evidence 7. Enforceability and redress 8. Compellability 9. Explainability 10. Constructiveness 11. Conduct 12. Learning organisation AP4AI is a project coordinated by Europol and CENTRIC with the support of Eurojust, EUAA and CEPOL. Europol (European Union Agency for Law Enforcement Cooperation) CENTRIC (Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research)
Status	Proposed

RP Specific ID	REQ-50.01.02 (SHR)
Source	01: mail Subject - AP4AI principles - on Tue 7-12-2022 5:15 PM
Rationale	01: N/A
Keywords	AP4AI; blueprint; security; accountability principles; regulatory; framework
Priority	Must have
Target / Sub target	CRRS
Internal owner	eu-LISA Business Analyst
External owner	[redacted] (Lot1 Consultant)
Final approver	eu-LISA Programme Manager
Acceptance criteria guideline	All the lower-level related requirements (e.g. SHR and/or FR and/or NFR/SEC connected in the hierarchy tree structure) have been realised and validated/tested, AND the description of this stakeholder requirement is fulfilled and it becomes 'true'.
Sector	(BRMS) Business Relations Management Sector

10 Appendix IV – Methodology for the identification of Business Use Cases, Requirements and AI Tools

10.1 Literature processing methodology

To drive literature analysis and processing, Lot1 team's work is driven by Qualitative Data Analysis method (Miles and Huberman, 1994). At a high-level, the process looks to process the information in an organised way though data reduction, data display and conclusion drawing and verification:

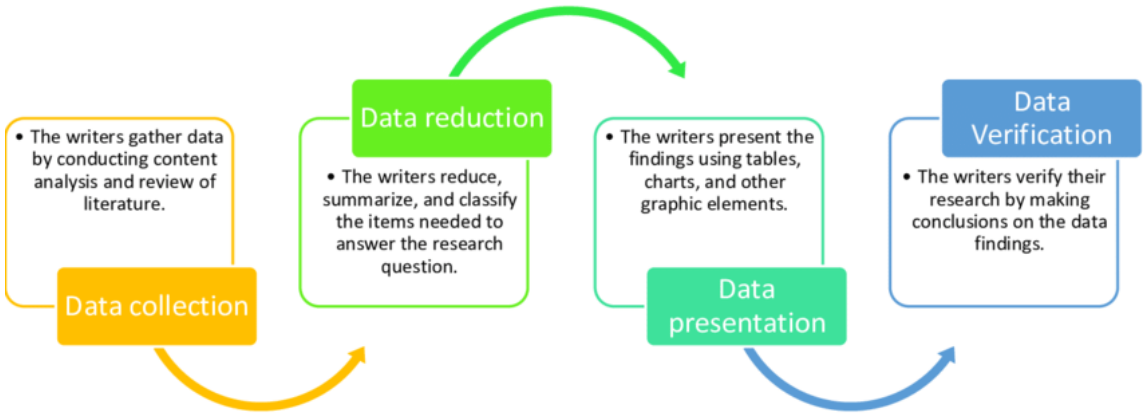


Figure 55. Literature processing methodology

10.2 Information gathering Methodology

The following data collection methods is used during the different phases:

One-on-One Interviews: also known as face-to-face interviews, allows to extract information relevant to our analysis from stakeholders. The discussion starts by informing the stakeholder about the activity and the objective helping to bring focus to the interview. A list of questions is shared before the interview, A benefit of this data collection methods for stakeholders is that the stakeholder is not under any pressure from other participants.

Focus Group Discussion: invitation to stakeholders to express their opinion on the objective. The number of people in a focus-group typically varies from 4 to 12 participants. The session chair ensures that no participant dictates the discussion. Discussions should not be based on preconceived notions to be successful. Active listening: the session should be used to better understand the business and its needs. Participants should not only agree or disagree, but also express opinions.

Collaborative Workshop Mode: the most engaging manner to collect and analyse data with participation from all stakeholders. Depending on the needs, the sessions may be spread over a few sessions, alternatively it could be concluded in one session. The stakeholder participation is not only used to gather data, but also draw conclusions. This enables a greater degree of engagement from stakeholders and helps in setting expectations. It may require multiple moderators.

Surveys: an economical method of data collection from a large audience or when low availability. Consists of a set of questions that are answered by stakeholders (typically by an Internet application) in each timeframe in which stakeholders need to complete the survey. Survey is appropriate for closed-ended questions.

10.3 Business Use Cases Definition Methodology

For Business Use Case elicitation, provided documentation was reviewed (see 1.8), workshops were scheduled for elicitation of needs, the meeting minutes were shared and a report on BUCs candidates was presented to all Stakeholders. The results of the workshops enabled the identification of use cases candidates. Additional Workshops for Requirements Elicitation are planned to be executed during Phase 2.

For the final document, the Unified Modelling Language (UML) framework was selected as a tool to define business functions and system functionalities required by the system.

The business use case model divides all use cases into following categories:

Business Use Cases,

Internal Use Case,
External Use Cases.

For the diagrams readability the colour codding has been applied in the models. Colour codding is described in the table below.

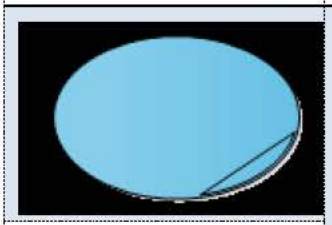
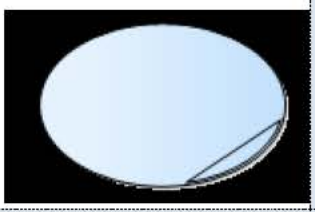
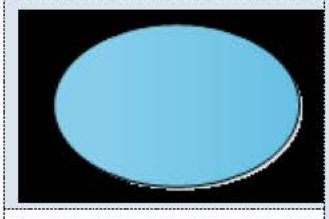
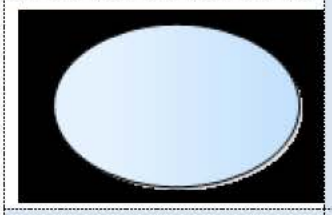
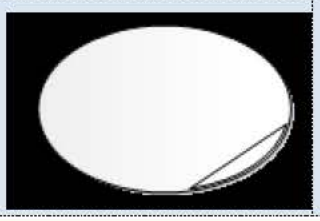
Symbol and colour code	Category	Applicability
	Business Use Case	The main subject of the diagram.
		Other Business Use Cases on the diagram.
	Internal Use Case	The main subject of the diagram.
		Other Internal Use Cases on the diagram.
	External Use Case	Functionality implemented in external system.

Table 38. Colour codes used in use cases diagrams

A **Business Use Case** (BUC) describes the general interaction between a business system and the users/actors of that system to produce **business results**. An example of the BUC diagram can be found on the following figure:

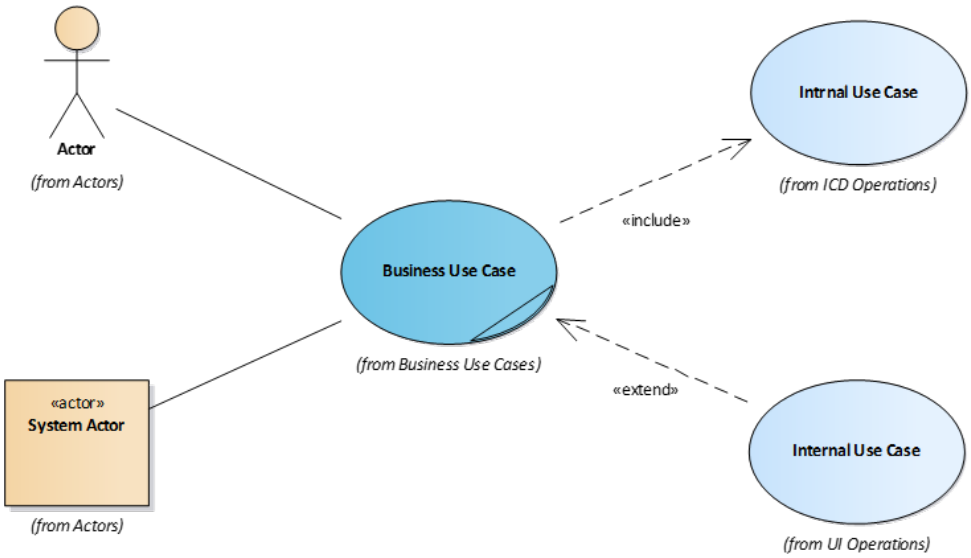


Figure 56. Business Use Case diagram example

The Business Use Case diagram contains a Business Use Case and related objects, the Actors that the Business Use Case is assigned to and the Internal Use Cases that shall be used to perform functions represented by the Business Use Case.

An **Internal Use Case** (Use Case, UC) defines the interactions between a role and an information system to achieve a **goal**. An Internal Use Case used in this document is close to the UML System Use Case definition. An example of the Internal Use Case diagram, the element of the Use Case model, can be found on the following figure:

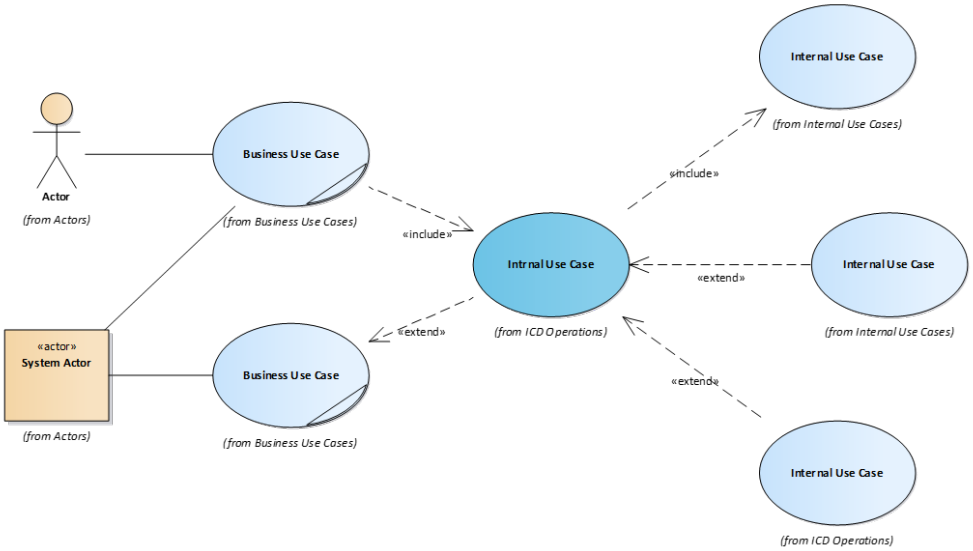


Figure 57. Internal Use Case diagram example

The diagram contains an Internal Use Case as a main subject (blue) and related objects: actors, related Business Use Cases and other Internal Use Cases used in the context of the main Use Case.

To keep the business context and make a ground for traceability, the Internal Use Case diagrams contain the Business Use Cases. Actors' assignment is indirect, via Business Use Cases. Other Internal use cases relate to standard relationships between use case showing elements required to achieve the use case goal.

From the point of view of access, the Internal Use Cases can be divided into the following categories:

- Available via user interface – most similar to classic UML Use Cases describing functionality available to End User that can contain several steps of interaction.
- System operation – available via ICD, functionality dedicated to use by other systems where interaction is limited to one request and response (which can be optional for asynchronous operations).
- Internal operations – structure used for potential functional reusability or functional extensions. It is used also in the cases where it supports model clarity.

The **External Use Case** defines functionality external to the system. It is used to specify functionality which takes place in course of Internal Use Cases activities or specifies the template (generalisation of reusable elements with other systems).

The simplified description is used to describe such Use Cases as the detailed implementation is out of the scope of the Contractor. The detailed specification for such functionalities shall be created in the other system's Interface Control Documents, considering scope of data downloaded from systems into CRRS.

Hence, the key difference between Internal and External Use Case is that Internal Use Cases concern the interaction only between internal roles or systems in the organization, while External Use Cases describe any interconnections outside of the system.

Use Cases are described with the following information:

Description element	Meaning	Business Use Case	Internal Use Case	External Use Case
Diagram	Use Case diagram.	Mandatory	Mandatory	Mandatory
Use Case name	Name of the Use Case. When Use Case is referred to from other sections this name is used.	Mandatory	Mandatory	Mandatory
Description	Describes the purpose and the objective of the Use Case. This part also refers to the legal base where relevant. The further implementation of the use case shall consider CRRS Technical specifications and requirements also, even if not mentioned in this document.	Mandatory	Mandatory	Mandatory
Business actor	Lists Use Case actors.	Mandatory	Mandatory NA for internal operations	Mandatory

Description element	Meaning	Business Use Case	Internal Use Case	External Use Case
Category	Indicates whether use case is: - Business Use Case. - In case of the Internal Use Case indication of its category: Available via user interface, system operation or internal operation. - External Use Case. - Operation accessible via user interface.	Mandatory	Mandatory	Mandatory
Inputs and outputs	Define information required to perform Use Case as well as output data or description of the results.	Mandatory	Mandatory NA for internal operations	Mandatory
Additional requirements	Lists Use Case requirements not covered or not fully covered by Use Case description nor Use Case Scenarios. The further implementation of the use case shall consider CRRS Technical specifications and requirements also, even if not mentioned in this document.	Optional	Optional	Optional
Scenarios	Main and possible alternate or exception scenarios which describe how the business goals can be achieved by an actor with use of available system functionality. The exception handling description covers business related errors only. Technical exception handling is not described in this document.	Mandatory	Optional	Optional
Scenarios diagram	UML activity diagrams presenting above scenarios.	Mandatory	Optional	Optional

Table 39. Use Case description structure

10.4 Requirements Methodology

10.4.1 Categories of Requirements

The following structure is proposed by the Guide to the Business Analysis Body of Knowledge (BABOK) v3, where the following types of deliverables are distinguished:

- Business Requirements (BR) - high level requirements capturing business motivation. They answer the question "why" Solution Capabilities are needed. Usually, they are derived from the Regulations. The BRs

will be translated into the solution requirements, for their implementation into the CRRS solution.

- Stakeholders Requirements (SHR) - requirements that identify the needs of stakeholders to achieve the business requirements. They describe "what" is needed by a specific stakeholder. These derive from the Commission Implementing Regulations and Commission Implementing Decisions related to the CRRS, ETIAS and revised VIS, other sources driven by documentation (see chapter 1.8.2) and workshops meeting minutes.

NOTE: Functional Requirements (FR) and Non-Functional Requirements (NFR) are outside the scope of this study.

10.4.2 Elicitation Process

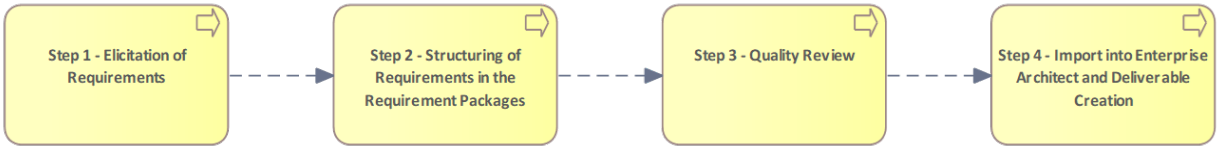


Figure 58. High-level view of the requirements elicitation process

10.4.3 Step 1 - Elicitation of Requirements

The requirements were elicited from the sources agreed with and provided by eu-LISA (See 1.8). The requirements are drafted in editable Excel files (securely stored in Lot1 SharePoint), which replicated the structure for requirements described in the previous section and enabled collaborative fast editing and import into Sparx Enterprise Architect (EA).

The high-level principles of the elicitation are outlined in the following paragraphs.

Business Requirements (BR):

Main Sources: legislation (Primary regulations).
Wording: as close as possible to the regulation and aligned with a glossary. The article numbers are included under attribute "Rationale" and avoided in the BR description.
Notes:
Each requirement shall be complete and exhaustive.
As much as possible, introducing parent-child relationship among BRs should be avoided.

Stakeholders Requirements (SHR):

- Main Sources: Commission delegated acts and implementing decisions, sources driven by documentation and workshops meeting minutes.
- Wording: "the solution xxx shall VERB for/to End-User Profiles / Authority / National System" e.g. "The System xxx shall make available fuzzy alphanumeric searches to the Immigration authorities". The perspective should always be from the system perspective, how system meets the business requirement.
- Notes: As much as possible, introducing parent-child relationship among SHR should be avoided. Every SHR must be linked to a BR

10.4.4 Step 2 - Structuring of Requirements in the Requirement Packages

To highlight the different views presented in the previous section, all requirements are presented in a hierarchy which is consistent with the single requirements package. For requirements that occur frequently, it is possible to link them between different requirements packages. The below diagram shows how a unique ID for each requirement could be derived and presented according to the hierarchy for each requirements package.

Figure 59. Requirement Hierarchy Sample

10.4.5 Step 3 - Quality Review

To ensure elicited requirements are in line with eu-LISA expectations, the following actions will take place:

Requirement's "kick-off" meetings with eu-LISA representatives.
Internal and ongoing QA (Quality Audit) on specifications for some of the requirements (e.g., legal or architecture expertise).
Ad-hoc communications with eu-LISA project team and extended team to solve any open questions.
Regular "[AI in CRRS] - Questions regarding requirements" emails with batches of questions to clarify open points.
Incremental Model applied to delivery as a process where requirements are divided into multiple modules for the requirements revision cycle.

10.4.6 Step 4 - Import into Enterprise Architect and Deliverable Creation

Finally, the collected requirements will be imported from the MS Excel files generated in Step 1 into Enterprise Architect (EA) Sparx, which was chosen as the main tool for requirements management.

For the review, Lot1 prepares an extraction from EA to MS Word files which are augmented by more descriptive introductory and explanatory texts, as well as diagrams that outline the hierarchy of requirements within the context of the requirements package.

10.4.7 Definitions of Attributes and Property Types

Property	Description
RP specific ID	Hierarchical requirement identification number, indicating the link to defined (Business) Use Cases and structure within requirements. The convention is: For BR: BUC#.sequential# (type) e.g., 8.1 (BR) For SHR: BUC#.linked_BR#.sequential# (type) e.g. 8.2.1 (SHR)
Requirement ID (Unique ID)	Unique requirement identification number. Convention is the following: IRS-REQ-sequential# (5 numbers) At this stage, the Requirement ID corresponds to the original ID from the Requirement Matrix present in the Reopening of Competition package to ensure backward traceability.
Requirement Name	Unique and meaningful short name used for initial identification of requirement.
Detailed Description of the Requirement	The main sentence where a requirement is formulated. BR: as close as possible to the regulation and aligned with a dictionary. SHR: *the System/component shall VERB for/to End User Profiles / Authority / National System* . The perspective should always be from the Systems/component perspective;

Property	Description
Status of a Requirement	A status of the requirement based on its stage in the lifecycle. The possible statuses are: proposed, verified, accepted, postponed, cancelled, implemented.
Source of a Requirement	Indicates where the requirement comes from. It shall be a document (a document's name) or a person (person's name, position, the date of a specific meeting/interview) or a group (e.g., Advisory Group (AG)).
Rationale	Justification/reasoning for the requirement. In case it is rejected, the reason for the rejection should be given here.
Keywords	Defines additional keywords facilitating the filtering of requirements.
Priority of a Requirement	Defines the priority of the requirement as per MoSCoW method: Must Have, Should Have, Could Have, Will not Have. Must Have: Describes a requirement that must be satisfied in the final solution for it to be a success. Should Have: Represents a high-priority item that should be included in the solution if it is possible. This is often a critical requirement but one which can be satisfied in other ways if strictly necessary. Could Have: Are defined as wanted or desirable but less important; and with less impact if left out (compared with a Should Have). Will not Have: Describes a requirement that the project team decided not to deliver (as part of this timeframe).
Target and Sub Target	Identifies the architectural element (e.g., Software (SW), Hardware (HW), infrastructure) where the requirement (functionality) should be applied. refers to the Target/Sub Target section for the possible values.
eu-LISA Requirement's Owner	A person within eu-LISA responsible for a formulation of a requirement.
External Requirement Owner	A person external from eu-LISA (i.e., from Contractor performing the elicitation), responsible for a formulation of a requirement.
Requirement's Final Approver	A person within eu-LISA responsible for a requirement approval.
Acceptance Criteria Guideline of a Requirement	Defines the criteria for acceptance guideline of the requirement as being fulfilled. It may include: the description of the state that must be achieved. It can be accompanied by a measure or/and target value, expressed in a value type based on Standard Index values, so it is possible to measure that the requirement has been met. In case of BR's and SHR's it will be a truth value.
Sector (eu-LISA)	Identifies eu-LISA sector responsible for the requirement.
Initiation Date	The date when the requirement is raised.
Requirement Type	BR, SHR.

Table 40. Definitions of Attributes and Property Types

10.5 Methodology for Identifying Artificial Intelligence Tools

In this project, we execute a series of actions based on **DMADV methodology (Define, Measure, Analyse, Design, Verify)** to determine the AI technologies that will support the identified Business Use Cases and Requirements. The individual steps of this methodology are summarized as follows:

- **Problem Definition:** The goal of the *problem definition* phase is to identify the process/service that requires enhancement or improvement. This is done in two instances: first conducting a high-level assessment of the current process (i.e., how risk profiling, screening rules and risk indicators are currently being handled);

and second how enabling an AI strategy will enhance/optimise the already existing process.

- **Measure:** In this step, the requirements that were gathered during the elicitation process will be used to assess the current technological framework within eu-LISA. Additionally, we will identify the interaction between CRRS, ETIAS and Recast VIS, and exchange of information between the various CBSs. Key milestone of the second step of the methodology is to discuss and agree with eu-LISA's stakeholders on the approach and expectations, and further elaborate on questions related to the activities during Phase 2 workshops.
- **Analysis:** The third step corresponds to the detailed investigation that will be implemented to determine the required information so that the AI Use Cases will produce actionable outcomes. This step is also implemented during Phase 1 of the study, to provide proactive and consistent outcomes for the BUCs in the context of AI and ML-based technologies.
- **Design:** The fourth step of the proposed methodology, refers to the study and the further assessment of the list of the identified AI tools/ frameworks. Moreover, this step will address the feedback received from the Interim Report and will adapt accordingly to the line of work.
- **Verification:** The fifth and last step of this analysis is the delivery and composition of the Final Report (i.e., Applicable AI technologies and Impact Assessment) which will describe the list of the proposed AI tools/frameworks addressing the demands that are identified on Phase 2 of the study, along with the received comments during the review process of the Interim Report.

Figure 6 illustrates a characteristic example of the activities that will be adopted in the context of enabling the BUCs with novel AI formulations. Specifically, the initial step of our assessment corresponds to the proper problem definition. The subsequent step refers to the explicit analysis of the problem as defined earlier and its break-down into different fundamental components. Moreover, the next step highlights the identification of the viable solutions that could facilitate on our analysis. As a subsequent step, we consider the selection of the optimal solution considering an AI approach for each different BUC. Finally, the last step is to conclude in a Case Study Generation, i.e., provide a well-defined analysis with the extracted business use cases and our suggestions regarding AI tools, technologies and solutions.

Figure 60. AI Methodology Identification Scenario

Before analysing our AI approach, it is essential to emphasize that Machine Learning (ML) is a key contributor to all data analytics processes. Data analytics processes may be performed either manually or automatically by using ML. Examples of such process are presented below:

Establish the minimum DQ Standards: For achieving the objectives CBS/IOD, it is important to determine the minimum DQ Standards.

Gather the required data: The process is initiated by collecting the appropriate data from one or more data sources. Usually in this type of procedures, the ingestion is not implemented directly, but requires suitable tools for ingestion depending on the platform in place.

Identification of data trends: Data trends and patterns shall be recognised and noticed using advanced machine learning algorithms analysing historical data and make predictions or conclusions regarding future data.

Improve ETL processes: An enhancement on the ETL processes will be performed by finetune the transformation Rules (CRRS-BO-11) and CRRS data structures (CRRS-BO-5)

Establish excessive storage consumption

Figure 8 depicts part of the flow of data analytics process based on machine learning algorithms

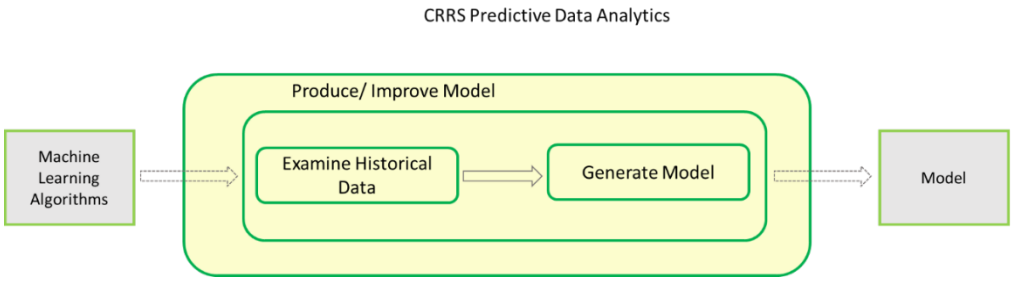


Figure 61. CRRS Predictive Data Analytics Scenario

To sum up, appropriate AI algorithms will be selected specifically to cover the organization's, stakeholders and subsequently CRRS and ETIAS needs regarding screening rules/ risk profiles/risk indicators and will further utilize historical information from various data sources (i.e., ETIAS, VIS, EES) and generate the respective ML models. After ML models have been produced, the next step of the process is the model validation/ evaluation phase, which is responsible for examining new data and generating predictions according to the already trained models. Since the current project, is a data-driven one, the Consortium adapts to the incoming data changes. For this purpose, models will be retrained to reflect the changes of the underlying environment. The parameters and variables that were initially used to train the models will not be changed; the updated models will adapt to current data, to result in more up-to-date and accurate predictions/ outputs. Model retraining is also highlighted on the Machine Learning Operations (MLOps) section of Chapter 5.

For the identification of the viable solutions for each BUC, Prescriptive and Predictive Analytics techniques are considered accordingly for the CRRS, Recast VIS and ETIAS data distributions. Additionally, Advanced Analytics techniques are investigated per BUC to provide added value in CRRS data.

In general, AI technologies can be categorized into three types, based on business capabilities:

- **Process Automation**: This category is dedicated to automating processes either physical or digital by using robotic process automation.
- **Cognitive Insight**: Discovering patterns from large data sets of diverse types and proceed into inferences. Usually, these kinds of systems are based on machine learning algorithms and are used to predict future behaviour.
- **Cognitive engagement**: Natural Language Process (NLP) algorithms shall be used for customer engagement. Examples of these systems are the chatbots and automated voice-controlled assistants (I.e., Alexa- Amazon, Google Assistant, Siri - Apple).

Among the most dominant approaches examined are: Deep Learning formulations, including Convolutional Neural Networks (CNNs) for identifying patterns in the input data, Recurrent Neural Networks (RNNs) and Long-Short Term Memory Networks (LSTMs) for time-series processing, Natural Language Processing (NLP), Text Recognition and Classification for processing large amounts of structured and semi-structured textual information, and Chatbot technologies.

Additionally, the AI methodology that corresponds to the various BUCs, relies on the basic principles of Machine Learning Operations (MLOps), by involving a set of processes and a sequence of steps implemented to deploy ML/AI models to production environments. Via this approach, we will be able to design and further create AI/ML models that are reliable, accurate, fast, and thus can be utilized by substantial number of concurrent users or applications. The key steps of the MLOPs are summarized in **Figure 8** including the definition of the ML/AI-based problems for the various BUCs, proceeding with the access and data preparation for model training, then with the architectural design of the ML solution, the development and deployment of the ML/AI models, and finally with the iterative process of monitoring, optimising and maintaining the performance of those models. MLOPs are thoroughly explained in Chapter 5 of the study.

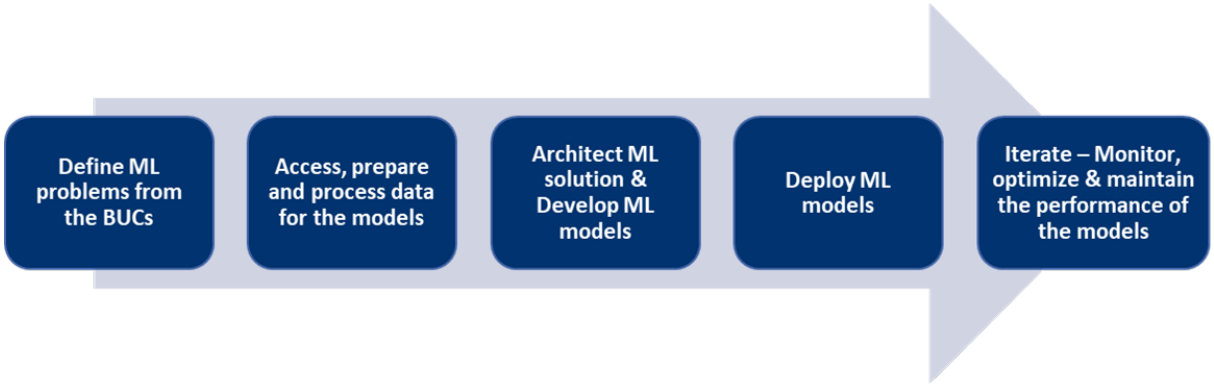


Figure 62. Machine Learning Lifecycle

Therefore, in the next steps of our Analysis and particularly in Phase 3, we investigate in further detail all the above-mentioned action points and propose the optimal ML/AI solutions and tools for the individual BUCs. Concluding, within the next Chapters of this report, we provide the complete analysis of the AI methodologies that are exploited in this study applying on the individual BUCs.

11 Terminology

Only abbreviations, acronyms and definitions in this document that are not in the Common Glossary are presented.

11.1 Abbreviations and Acronyms

TERM	FULL NAME
AI	Artificial Intelligence
AP4AI	Accountability Principles for AI
BA	Business Analyst
BABOK	Business Analysis Body Of Knowledge
BR	Business Requirements
BUC/UC	Business Use Case
CBS	Core Business System
CIR	Common Identity Repository
COM	The European Commission
CRRS	Central Repository for Reporting and Statistics
DMADV	Define, Measure, Analyse, Design, Verify (Six Sigma framework)
DoE	DevOps Engineer
DS	Data Scientist
EA	Enterprise Architect
EASO	European Union Agency for Asylum
EBCGA	European Border and Coast Guard Agency (a.k.a. FRONTEX)
ECDC	European Centre for Disease Prevention and Control
ECRIS-TCN	European Criminal Records Information System for Third-Country Nationals
EDPS	European Data Protection Supervisor
EES	Entry Exit System
ESP	European Search Portal
ETIAS	European Travel Information and Authorisation System
ETIAS CS	European Travel Information and Authorisation System Central System
ETIAS CU	European Travel Information and Authorisation System Central Unit
ELT	Extract, Load, Transform
ETL	Extract, Transform, Load
EU	European Union
FTE	Full Time Equivalent
eu-LISA	European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice
FQP	Framework Quality Plan
FR	Functional Requirement
FRA	European Union Agency for Fundamental Rights
HW	Hardware
ICD	Interface Control Document
Id	Identification

TERM	FULL NAME
ID	Interface Designer
IO	Interoperability
LTV	Limited Territorial Validity
MID	Multiple-identity detector
ML	Machine Learning
MVP	Minimum Viable Product
MoSCoW	Must Have, Should Have, Could Have, Will not Have (Agile prioritisation technique)
MS	Member State
NLI	Natural Language Interface
NLP	Natural Language Processing
NU	National Unit
NFR	Non-Functional Requirement
OS	Open Source
PDF	Portable Document Format
PD	Project Director
PM	Project Manager
PoC	Proof of Concept
QA	Quality Assurance
RP	Requirements Packages
SA	Solution Architect
SHR	Stakeholders Requirements
SE	System Engineer
SEC	Security Requirements
SIS	Schengen Information System
SPoC	Single Point of Contact
Statistical reports	Means an organised collection of statistical data, produced by the central repository in an automated manner according to a set of pre-established rules and stored in the central repository
SW	Software
TA	Travel Application
TE	Tester Expert
TV	Travel Visa
TCN	Third-Country National
UML	Unified Modelling Language
VIS	Visa Information System
WHO	World Health Organization
XML	Extensible Markup Language

Table 41. Abbreviations and Acronyms