EXPLORATORY RESEARCH

sesar JOINT UNDERTAKING

Guidelines on Decision Intelligence in Air Traffic Management

Deliverable ID:	**D2.3**
Dissemination Level:	**PU**
Project Acronym:	**SafeOPS**
Grant:	**892919**
Call:	**H2020-SESAR-2019-2**
Topic:	**SESAR-ER4-06-2019**
Consortium Coordinator:	**TUM**
Edition date:	**31 October 2022**
Edition:	**00.01.00**
Template Edition:	**02.00.03**

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & Approval

Authors of the document

Name / Beneficiary	Position / Title	Date
Lukas Beller / TUM	Project Coordinator / M.Sc.	31.10.2022
Elizabeth Humm / DBL	Project Member / Ph.D.	25.10.2022
Carlo Abate / DBL	WP3 Leader / Ph.D.	25.10.2022
Pablo Hernandez / INX	WP4 Leader / M.Sc.	25.10.2022
Philipp Kurth / DFL	Project Member / ATCO	26.10.2022

Reviewers internal to the project

Name / Beneficiary	Position / Title	Date
Lukas Beller / TUM	Project Coordinator / M.Sc.	31.10.2022
Elizabeth Humm / DBL	Project Member / Ph.D.	28.10.2022
Carlo Abate / DBL	WP3 Leader / Ph.D.	28.10.2022
Phillip Koppitz / TUM	Project Member / M.Sc.	31.10.2022
Philipp Kurth / DFL	Project Member / ATCO	28.10.2022

Reviewers external to the project

Name / Beneficiary	Position / Title	Date

Approved for submission to the SJU By

Name / Beneficiary	Position / Title	Date
Lukas Beller / TUM	Project Coordinator / M.Sc.	31.10.2022

Rejected By - Representatives of beneficiaries involved in the project

Name and/or Beneficiary	Position / Title	Date

Document History

Edition	Date	Status	Name / Beneficiary	Justification
00.00.01	24.06.2022	Initial Scope and Outline	Lukas Beller / TUM	-
00.00.02	12.10.2022	Draft Content in InGrid	Lukas Beller / TUM	-

00.00.03	19.10.2022	Reviewed Content on Predictive Layer	Pablo Hernandez / INX	-
00.00.04	25.10.2022	Reviewed Content on Risk Framework	Elizabeth Humm / DBL	
00.00.05	27.10.2022	Reviewed Content on Operational Layer	Lukas Beller / TUM	-
00.01.00	31.10.2022	Implementing Content in Template and Formatting	Lukas Beller / TUM	-

SafeOPS

FROM PREDICTION TO DECISION SUPPORT - STRENGTHENING SAFE AND SCALABLE ATM SERVICES THROUGH AUTOMATED RISK ANALYTICS BASED ON OPERATIONAL DATA FROM AVIATION STAKEHOLDERS

This deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 892919 under European Union's Horizon 2020 research and innovation programme.

Abstract

This deliverable is the final technical deliverable of the SafeOPS project. It summarizes the processes and methodologies, applied in the three technical work packages within the project and compares the performed work against the EASA Guidance on AI Application. Thereby, the deliverable summarizes the open tasks for further maturity and lessons learned. Furthermore, we provide additional and complementary guidelines, relevant for decision intelligence concepts in Air Traffic Management, distilled from the lessons learned.

Table of Contents

List of Tables

List of Figures

EASA Objectives

SafeOPS Guidelines

EUROPEAN PARTNERSHIP

Co-funded by the European Union

EUROPEAN PARTNERSHIP

Co-funded by the European Union

1 Introduction

The next generation Air Traffic Management (ATM) systems are pushed more and more towards digitization, fueled by the demands to increase capacity and cost-efficiency while also increasing the already high safety and resilience levels. SafeOPS proposes and investigates a decision support tool, which provides Air Traffic Controllers with predictive risk information on the likelihood of approaches to perform a go-around, in real time. The underlying idea is to enable Air Traffic Controllers to incorporate predictive analytics in pre-planning during departure and approach handling, improving the decision-making in terms of capacity as well as safety and resilience.

Resilience is defined differently in various contexts. From Oxford Advanced Learner's Dictionary [1], one intuitive definition in a personal and human context is "the ability of people (...) to feel better quickly after something unpleasant, such as shock, injury, etc. happened". In general, resilience describes the ability to absorb, adapt or recover from rare or unpredictable events and disturbances. This definition is not limited to humans, but can also be applied to systems, organizations, or operations. In SafeOPS, we focus on the engineering perspective of resilience, which describes the ability and rate at which a system returns to an equilibrium after a disturbance or disruption.

Disruptions affect a system or organization in normal operation. They cause discontinuity, confusion, disorder, or displacement and can take various forms. Madni and Jackson [2] p. 2 categorized major disruptions as originating from **operational contingencies**, natural disasters, political instability, or economic events. Resilient systems can react and adapt to these disruptions. While it is possible to protect a system against known or foreseeable events by analyzing historical data/events, it is much harder to anticipate and counter rare or even completely unknown disruptions. This implies that systems need adequate safety margins to cope with uncertainty [2]. In [2], the authors conclude that **humans are more capable** of detecting and **handling unpredicted situations than machines**. Thus, **humans are considered as net assets** for resilience in systems like Air Traffic Management.

As of today, Air Traffic Controllers recognize the onset of go-arounds through pilots' communication and from observing the aircraft visually or via radar. Go-arounds are a standard and well-established flight procedure, for flight crews as well as for Air Traffic Controllers. Within the states of ECAC, they occur with an average rate of 3 per 1000 approaches [3]. Despite the relatively low likelihood of a go-around to occur, Air Traffic Controllers have strategies for handling them, readily available from initial training. These strategies are of reactive nature, meaning they become active once the Air traffic Controller identifies the ongoing go-around. But especially in high traffic congestions, the Controller might realize that a go-around is ongoing after a departure has been cleared for take-off or is airborne already. In these situations, handling a go-around becomes complex, since knock-on effects like separation infringement or wake turbulence challenges with preceding departures can arise. This is even more relevant, if the standard missed approach procedure conflicts with the departure route of the preceding aircraft. Air Traffic Controllers are trained for such situations, nevertheless, resolving such situations, while maintaining safe separation, increases the workload of the Air Traffic Controller, as well as the Flight Crew, since additional clearances as well as additional coordinative actions to adjacent sectors become necessary at the same time. The controller has to prioritize immediately the following actions and tasks. These situations serve as good examples for the resilience of the ATM system and the role of human beings in the system, in the sense that a rare event and the possible safety relevant knock-on effects are mitigated through human actions.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

It must be emphasized that the described go-around scenarios, including traffic congestion and a possible conflict between missed approach and departure routes, are only a subset of all go-around scenarios. Especially in times of medium to low traffic volume, pilots most likely can fly briefed standard missed approach procedures without any safety relevant knock-on effects occurring. In these cases, the existing Safety-I based procedures have proven and will continue to prove adequate safety barriers with Air Traffic Controllers providing the necessary resilience in the Air Traffic Management System.

However, fueled by Covid-19, a general debate on a "Resilience versus Efficiency" trade-off emerged [4], [5]. In [6], Vardi transfers the lessons learned from the Covid-19 pandemic to the computing domain (The content of the article is also available as online lecture on youtube). Thereby he partly pounces on the discussion in [7], which describes the Covid-19 related impact on US businesses, followed-up with the question whether this domain, streamlined towards efficiency and thus lacking the adaptive capacity (required for resilience), has become vulnerable to disruptions.

This debate is also relevant for Air Traffic Management. The European Air Traffic Management Master Plan's ambitions [8] envisions a 60% increase in network throughput of IFR flights by 2035, compared to 2012. Simultaneously, the ambitions include an increase in safety levels by 100%. Following the "Resilience versus Efficiency" debate, these two goals could prove challenging to conciliate. The first contributing factor for complex go-around situations, the conflicting procedures, roots in environmental conditions around an airport, like terrain or noise abatement restrictions, and is independent from the stated ambitions. High traffic load situations in the ATM network, however, will increase, also making the scenario for complex go-arounds more frequent, which is the second contributing factor. Therefore, complex go-around situations serve as good scenarios to investigate how decision intelligence could provide a benefit for safety and resilience in Air Traffic Management.

Decision intelligence is an engineering discipline providing a framework which incorporates (predictive) data science in decision-making processes [9]. It aims to support, augment or automate decisions, by providing additional insights gained from data, which were not available in the prior decision-making process. In SafeOPS, the decision intelligence framework was used to investigate the decision-making processes of Air Traffic Controllers under the premise that data-driven go-around predictions are provided to them. The idea when providing **predictive** risk information in this scenario is, that Air Traffic Controllers are enabled to use proactive tactics instead of the reactive tactics described above, to avoid the described knock-on effects possibly triggered by a go-around. The proactive approach introduces a Safety-II based feature, which complements the already existing, Safety-I related, reactive tactics. Furthermore, the proactive solution benefits the resilience of the ATM system by providing the ATCOs with more time and better information for the necessary coordinative actions, which have to be taken in the event of a go-around.

Within this deliverable we summarize the lessons learnt from the SafeOPS project on decision intelligence in Air Traffic Management. Thereby, we focus on an operational, predictive and risk oriented perspective, covering the systems engineering, human factors integration, and machine learning processes. The HMI perspective, including a set of 16 guidelines, was already presented in Deliverable 3.3 (D3.3) Section 4 / Appendix B [10].

In December 2021, in the middle of the SafeOPS project and well after the project proposal was written, EASA published an extensive First usable guidance for Level 1 machine learning applications [11]. The EASA document contains a lot of relevant guidance on the development of AI based systems with the level of automation of the decision support tool worked on in SafeOPS. It was therefore not the goal of this task to reinvent the available guidance. Rather, we aim to:

EUROPEAN PARTNERSHIP

Co-funded by the European Union

- analyse to what degree SafeOPS followed the EASA guidance material,
- document where SafeOPS deviated from the EASA guidance material, and
- report, where additional guidelines can be concluded from SafeOPS.

In the following, we will review our work with focus on the lessons learned. This will be done for each technical work package:

- Operational Layer,
- Risk Framework and
- Predictive Layer.

From this we will derive guidelines and try to support these with references to additional guidance material.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

2 Operational Layer

The operational related aspect of the decision intelligence concept focuses on:

- the operational procedure design, which includes the investigation of state-of-the-art approaches and go-around handling as well as use cases for a decision support concept,
- integration of ML predictions in ATM, including the integration of stochastic information into the operation,
- the data selection and acquisition, based on the requirements concluded from use cases,
- and the impact assessment methodology, to measure the concept's influence on safety, resilience, and capacity of ATM, especially in the go-around scenario.

Deliverable 2.1 (D2.1) [12] covered the first three focal points, Deliverable 2.2 (D2.2) the latter. Due to the project proposal's deviation from the SESAR project organization framework, what was initially planned to be reported in D2.2 is split into a non-contractual Experimental Plan deliverable and D2.2, which was written in the form of a Validation Report.

For a decision intelligence solution, it is crucial to understand the decision-making processes in the targeted scenario(s). This includes state-of-the-art decision making, as well as triggering events for the relevant decisions. Furthermore, it is crucial to work out the envisioned decision-making processes, including the provision of the machine learning solution's predictive information. Under this premise, SafeOPS composed a methodological approach that combines methods from agile development, resilience engineering [13] and systems engineering. The methodology is described in deliverable 2.1 (D2.1) - Section 2. In essence, especially the Scenario and Use Case definition (comparable to the SJU's reference and solution scenario definition in e.g. SESAR Experimental Approach Guidance ER), including sequence diagrams to document the flow of actions and decision making proved valuable throughout the project. To define these Scenarios and Use Cases, SafeOPS organized recurring workshops during the initial phase of the projects, including the foreseeable end users of the concept. The workshop planning oriented along the resilience engineering methods *Work as Done* and *System Thinking* [14]. The *Work as Done* principle especially helped the development team in understanding an ATCO's behavior in different situations, especially his/her strategies to mitigate risk. The *System Thinking* principles, especially through the focus on *Local Rationality, Demand and Pressure and Resources and Constraints,* supported the Use Case definitions, indicating the limits of current procedures and indicating where support in decision-making can provide positive impacts in the defined scenarios. This methodology also provides a good approach to generate the requested items for **Characterization of the AI application Objectives** from the EASA Guidelines. Before diving into the AI/ML objectives, SafeOPS Guideline 1 should be considered.

SafeOPS Guideline 1

SafeOPS Guideline ID	Guideline
SO.GL.1	The applicant should identify possible conservative (non-AI/ML) solutions in their initial scoping of the concept

EUROPEAN PARTNERSHIP

Co-funded by the European Union

SafeOPS Guideline 1 arose from an early discussion, when developing the SafeOPS concept with end-users. Some end-users indicated that they would be happy if they only had a fast, reliable and deterministic real time go-around detection and indication, rather than a non-deterministic, time-ahead predictive information. The argument was that pilots, following the 'aviate, navigate, communicate' paradigm will communicate the go-around to the relevant Controller with a significant time delay. In case of bad visibility, the ATCO might not realize the ongoing go-around on the radar immediately and thus, relevant **reaction** time is lost. The proposed research topic of SafeOPS was however on risk prediction and also from the ongoing discussions in workshops we found that predictive go-around information can provide benefits over the go-around indication, however it should be stated that from an economic and regulatory perspective, the detection concept would be easier to implement. Regarding regulatory challenges, the upcoming discussion will focus on the guidance of the AI/ML related challenges of the SafeOPS concept.

2.1 Characterisation of the AI application Objectives

In the following, we measure the work done in SafeOPS against the relevant objectives from EASA's guidance and state additional guidelines applicable for a decision intelligence concept, which we derived from SafeOPS.

2.1.1 High-level task(s) and AI-based system definition

The first relevant objective is stated below.

EASA Objective 1 [11]

EASA Objective ID	Objective
Objective CO-01	The applicant should identify the list of end users that are intended to interact with the AI-based system, together with their roles, their responsibilities, and their expected expertise (including assumptions made on the level of training, qualification and skills).

The end-users were identified and because of the nature of this user group, e.g. ATCOs and pilots, then ‘their roles, their responsibilities and their expected expertise’ were assumed, due to the fact that they are professionally licensed to fulfil their role. Moreover, using a series of go-around scenario descriptors, ‘involved actors’ (operator roles) were clearly detailed for each scenario. The SafeOPS project did not consider the full range of users (beyond the end-user), for example all those who might be part of the system at some point, including decision makers, managers, trainers, maintainers, monitors, and inspectors. At this exploratory phase of the project, more focus was placed on developing the concept of the technological solution and the interplay of the operational user.

Additionally to EASA Objective 1, SafeOPS identified two complementing guidelines for the addressed decision intelligence idea in SafeOPS Guideline 2 and SafeOPS Guideline 3.

SafeOPS Guideline 2

SafeOPS Guideline ID	Guideline

Co-funded by the European Union

SO.GL.2	The applicant should identify the human capabilities, limitations, performance and behaviour of the end-user.

The SafeOPS Guideline 2 considerations are more related to *human characteristics*, as opposed to something that has been trained and rehearsed. This guideline is about identifying inherent human attributes that cannot be trained or up-skilled to meet the AI technology integration needs. This is particularly pertinent to AI as its integration into safety critical systems is relatively new and thus knowledge of the human behavior alongside AI is scarce. This guideline should be achieved through a paper exercise or basic lab exercise in which these considerations are researched or observed and reported, to expand knowledge and understanding about the newly developing concept and technology. The purpose of this is, to understand how the new concept impacts human performance and how to optimize the design for enhanced human performance. Furthermore, it is important to review methods for testing human performance when using the technology and to identify the potential human interactions, issues and risks relevant to the new concept. An example from SafeOPS could be the limitation of the operator's cognition, when comprehending the significance of the display of a probabilistic value.

SafeOPS Guideline 3

SafeOPS Guideline ID	Guideline
SO.GL.3	The applicant should determine the need for additional/modified training in interacting with AI/ML based systems.

Training should be specifically reviewed to encompass current understanding on the need to support humans in the integration of AI technologies, e.g. in:

- education of the user on what should be expected by the observable decision process of the model,
- what is appropriate decision support advice,
- what are trustworthy behaviors of the technology (in order to engender trust) and
- the possible consequence of making decisions based on the technology outputs (to avoid over-reliance on AI).

SafeOPS Guideline 3 is relevant in SafeOPS since the absence of a prediction for one class (Go-around Prediction) does not imply a prediction for the second class (Landing Prediction). With the SafeOPS concept, ATCOs have to be aware that decisions to give take-off clearances should not depend on the absence of a go-around prediction.

The second relevant objective for the operational layer is EASA Objective 2.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

EASA Objective 2 [11]

EASA Objective ID	Objective
Objective CO-02	For each end user, the applicant should identify which high-level task(s) are intended to be performed in interaction with the AI-based system.

EASA Objective 2 was considered by the SafeOPS project during the work on developing a risk framework. The high-level tasks associated with the go-around were identified, namely runway management and monitoring, separation monitoring, wake vortex monitoring and trajectory management. These tasks are at the level of the interaction between the human and the AI-based system and were identified as critical for the purpose of risk modelling of the go-around event. The identification of these tasks assisted in the understanding of how the technology could support them and further work was done looking at how the technology insertion would impact them. Moreover, exploration of the high-level tasks was also conducted in the work done in SafeOPS on the operational scenarios.

Also EASA Objective 3 is relevant for the operational layer.

EASA Objective 3 [11]

EASA Objective ID	Objective
Objective CO-03	The applicant should determine the AI-based system taking into account domain-specific definitions of 'system'.

The work conducted in SafeOPS was grounded on a systems engineering approach through the work conducted on generating Scenarios, Use Cases, User Stories and Requirements in the early parts of the project in deliverable 2.1 (D2.1) [12]. Furthermore, as this project considers a safety critical environment, the methodology of *resilience engineering* was also applied during the early stages of the project, in order to direct the concept description and design activities towards the consideration of safety, reliability and resilience.

The other aspects of the 'system' as defined in the EASA paper, were not considered in SafeOPS due to the fact that it is exploratory research. These are namely a fully engineered solution, training, personnel selection, procedures, organizational and administrative support, and would be fully realized as the concept matured through its development. Specifically, further tests and evaluation of an increasingly mature prototype, in low- and high-fidelity environments, would be necessary to refine the details of their associated requirements.

2.1.2 Characterisation of the AI application - Concept of operations for the AI application

EASA Objective 4 is in line with the SafeOPS methodology, but also with the general methods requested by the SJU from SESAR funded projects

Co-funded by the European Union

EASA Objective 4 [11]

EASA Objective ID	Objective
Objective CO-04	The applicant should define and document the ConOps for the AI-based system, including the task allocation pattern between the end user(s) and the AI based system. A focus should be put on the definition of the operational design domain (ODD) and on the capture of specific operational limitations and assumptions.

Much work has been done in SafeOPS to describe a user-centric, operational concept for the technological solution. Throughout the project, regular engagement with the user community was essential in ascertaining the operational requirements for the technology. From this activity a high-level description of operational scenarios, operating limitations and operating conditions were defined through user workshops. An initial ConOps is described in deliverable 2.2 (D2.2) ref.

Operator tasks and sub-tasks, including decision making responsibilities, were developed through the ConOps generating activities and through human factors and safety analysis. This work generated a greater knowledge and understanding of the limitations and assumptions of the technological system, and thus allowed the designers to establish the appropriate placement of the technology in terms of task allocation patterns between user and technology.

2.1.3 Characterisation of the AI application - Functional analysis of the AI-based system

EASA Objective 5 [11]

EASA Objective ID	Objective
Objective CO-05	The applicant should perform a functional analysis of the system.

The project defines several subsystems for the envisioned concept, containing AI and HMI. The AI subsystem is further divided into data acquisition, data storage, data pipeline and model training, whereas the remaining subsystems are not defined in greater detail within this scope of the project. An open task, in case of further investigation of this concept is to develop a detailed documentation of subsystems and interfaces, including the radar screen for visualization, radar/ADS-B antennas for information retrieval and the AI subsystem.

Additionally, from the human integration perspective, SafeOPS suggest SafeOPS Guideline 4

SafeOPS Guideline 4

SafeOPS Guideline ID	Guideline
SO.GL.4	The applicant should determine the technology characteristics that support the relevant human capabilities and limitations.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Guidelines for the design of systems and technologies to support safe and effective human interactions should be developed, based on human characteristics, capabilities, and limitations defined in CO-01. This is particularly noteworthy when the 'language of explainability' is provided by the system to the operator. The design team should ensure that the information is presented to the user in a way that considers the limits of human cognition and understanding, but also the operational environment, work demands and the time component.

2.1.4 Characterisation of the AI application - Classification of the AI application

In order to determine the relevant applicable safety measures for an AI/ML solution, EASA suggests different levels of AI, which consequently have to prove different levels of compliance with the objectives. The classification is defined in EASA Objective 6.

EASA Objective 6 [11]

EASA Objective ID	Objective
Objective CL-01	The applicant should classify the AI-based system, based on the levels presented in Table 1— EASA AI typology and definitions, with adequate justifications

Table 1: EASA AI typology and definitions is provided in the following for convenience.

Table 1: EASA AI typology and definitions [11]

EASA AI Roadmap AI Level	Function allocated to the system to contribute to the high-level task
Level 1A Human augmentation	Automation support to information acquisition
	Automation support to information analysis
Level 1B Human assistance	Automation support to decision-making
Level 2 Human-AI collaboration	Overseen and overridable automatic decision-making
	Overseen and overridable automatic action implementation
Level 3A More autonomous AI	Overridable automatic decision-making
	Overridable automatic action implementation
Level 3B Fully autonomous AI	Non-overridable automatic decision-making
	Non-overridable automatic action implementation

During the project, the SafeOPS tool was not classified according to the EASA guidelines, however the tool in this project is aimed at decision support and therefor Level 1B is suggested. This will be

EUROPEAN PARTNERSHIP

Co-funded by the European Union

important in case of further work on the concept, as this classification determines further objectives and how they must be applied.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

3 Risk Framework

The Risk Framework, described in deliverables D3.2 [15], D3.3 [10] in detail, articulates the risk associated with the integration of the SafeOPS concept (the go-around prediction) into the ATM system. In a three-step approach, SafeOPS firstly identified the operations, decisions and actions which were impacted by the presence of the SafeOPS tool (as part of the operational layers systems engineering approach). Secondly, we described and integrated these components into the Accident-Incident Model (AIM), a risk model class from Eurocontrol and SESAR Joint Undertaking. Thirdly, by describing how the individual elements of the model change after introducing the SafeOPS tool. The first step of this analysis identified at a high level the safety functions fulfilled by the ATCOs before and during the go-around maneuver, namely:

- Runway management,
- Traffic separation monitoring,
- Monitoring of the wake category, and
- Trajectory management.

With the Risk Framework, it was possible to effectively identify the base events that were impacted by the introduction of the predictive tool, which in many cases involved the lack of sufficient time to timely assess, and react to, the evolving situation. The analysis revealed that there were several improvements to the safety of the system, from the introduction of the SafeOPS tool. These improvements included **increased situational awareness** of Air Traffic Controllers, more time to get an accurate and complete picture of the traffic, and **more time to perform their tasks**. These improvements have a smoothing effect on operators' workload and thus results in a lower probability of human errors, an increased chance that a potential conflict is identified and a higher likelihood that effective plans are made to anticipate or resolve potentially hazardous situations. Although considered highly unlikely, the analysis also found a **small number of drawbacks**. These include the eliciting of unsafe behaviors, such as **issuing clearances based on a disproportionate level of confidence** that an inbound aircraft will definitely land; and also, the act of cancelling a take-off clearance resulting in an increased risk of runway excursion.

This Risk Framework also provides a good initial approach to generate the requested items for **Safety assessment of ML applications** from the EASA Guidelines [11].

3.1 Safety assessment of ML applications

In the following, we measure the work done in the risk framework against the relevant objectives from EASA's guidance [11] and state additional guidelines applicable for a decision intelligence concept, which we derived from the risk framework actions.

3.1.1 Safety assessment of ML applications - System safety assessment

The relevant objective from EASA, regarding safety assessments of ML solutions are presented in EASA Objective 7 and EASA Objective 8.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

EASA Objective 7 [11]

EASA Objective ID	Objective
Objective SA-01	The applicant should define metrics to evaluate the AI/ML constituent performance and reliability.

In the predictive layer of SafeOPS, metrics for the performance of the AI constituent have been defined and evaluated. These also served as bases for the discussions in the Risk Framework. For performance evaluation, precision and recall metrics are used. To determine an optimal threshold for the binary classification algorithm, also Receive Operating Characteristics and Precision-Recall curves were considered.

The reliability of the AI/ML tool was not investigated thoroughly at this stage of the project, but is an open task, in case of further research on the concept.

Additionally, from the human performance perspective, SafeOPS proposes SafeOPS Guideline 5.

SafeOPS Guideline 5

SafeOPS Guideline ID	Guideline
SO.GL.5	The applicant should define objective and subjective metrics that will signal human-system performance, effectiveness, and safety.

The metrics for SafeOPS Guideline 5 should examine human behavior whilst interacting with the technology and may include accuracy, response and completion times, error types and frequency, workload, situation awareness, user satisfaction and acceptance, usability, trust in the system, accurate understanding of the system´s actions, and appropriateness of the system´s generated decision advice.

Based on the metrics from SafeOPS Guideline 5 and EASA Objective 7, EASA Objective 8 clarified the safety assessment.

EASA Objective 8 [11]

EASA Objective ID	Objective
Objective SA-02	The applicant should perform a system safety assessment for all AI-based (sub)systems, identifying and addressing specificities introduced by AI/ML usage.

Regarding EASA Objective 8, a full-up system safety assessment was not conducted in SafeOPS, as the project was of an exploratory nature. However, for the model of a 'Safety Scoping & Change Assessment' [16], the preparatory process of identifying the main safety issues associated with a specific solution was used to conduct preliminary analyses on system safety. The application of this methodology was appropriate to the SafeOPS tool as it represents a *change* to the current system and in terms of safety and risk modelling, it is a change to a risk model that already exists. In this scoping

Co-funded by the European Union

assessment, the SafeOPS technology was considered as a concept in the current system of ATM, as such the AI/ML detailed constituent performance and reliability data wasn't taken into account in this modelling. In this analysis the SafeOPS technology was considered in a barrier-based quantitative model of risk assessment. In using this model, performance requirements and quantitative safety performance targets were explored and safety objectives defined. As the system matures and as more data becomes available on AI/ML constituent performance and reliability, a full safety assessment should be conducted.

Especially for AI/ML solutions, we propose a complementary guideline, which takes into account the uncertainty in the output of ML solutions.

SafeOPS Guideline 6

SafeOPS Guideline ID	Guideline
SO.GL.6	The applicant should explore the output space for all AI/ML (sub)systems, and the impact of possible outcomes on the operation / user.

AI/ML solutions can provide non-deterministic information. It is important to incorporate into the assessment, that the prediction is not certain to materialized. In case of binary classification as used in SafeOPS, false positive (a predicted go-around, which will not occur) and false negative (a not predicted go-around which will occur) predictions must also be considered, when assessing the impact of a tool. The precision and recall metrics allow to weight the different prediction cases against each other, according to their relative likelihood.

3.1.2 Information security considerations for ML applications

Regarding information security, EASA proposes the following objectives, summarized in EASA Objective 9

EASA Objective 9 [11]

EASA Objective ID	Objective
Objective IS-01	For each AI-based system and its data sets, the applicant should identify those information security risks with an impact on safety, identifying and addressing specific threats introduced by AI/ML usage.
Objective IS-02	The applicant to document a mitigation approach to address the identified AI/ML-specific security risk.

SafeOPS has at this early stage of the developments, not performed a security assessment of the concept. The aspects of both objectives should be addressed in case of further research on the concept, as they are relevant in terms of security criteria of the SESAR TRL2 assessment.

3.1.3 Ethics-based assessment

The EASA Objective 10 and EASA Objective 11 are posed, regarding ethics assessment of AI solutions.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

EASA Objective 10 [11]

EASA Objective ID	Objective
Objective ET-01	The applicant should perform an ethics-based trustworthiness assessment for any AI-based system developed using ML techniques or incorporating ML algorithms.

An ethics-based technology trustworthiness assessment was not conducted as it was not considered when scoping the SafeOPS project. It is recommended that the Trustworthy AI Assessment [17] list is applied to the SafeOPS projects in any future iteration.

EASA Objective 11 [11]

EASA Objective ID	Objective
Objective ET-02	The applicant should comply with national and EU data protection regulations (e.g. GDPR), i.e. involve their Data Protection Officer (DPO), consult with their National Data Protection Authority, etc..

SafeOPS has access to different data sources for the project. First, Flight Data Monitoring (FDM) data from airlines, for which a data protection agreement between data providers and users regulates the storage and usage, as well as the deidentification of the data. Additionally, publicly available data, such as ADS-B and METAR are used, for which an agreement of how to use the data exists.

3.1.4 AI Safety Risk Mitigation - AI SRM top-level objectives

EASA Objective 12 [11]

EASA Objective ID	Objective
Objective SRM-01	Once activities associated with all other building blocks are defined, the applicant should determine whether the coverage of the objectives associated with the explainability and learning assurance building blocks is sufficient or if an additional dedicated layer of protection, called hereafter safety risk mitigation (SRM), would be necessary to mitigate the residual risks to an acceptable level.
Objective SRM-02	The applicant should establish SRM means as identified in Objective SRM-01.

Analysis on SRM's wasn't conducted in SafeOPS as not all activities associated with the building blocks were defined at this stage of the project. However it is envisaged that the use of SRMs will be a valuable part of the system in its initial release to service, during the period in which increasing amounts of usability and safety data will become available.

Co-funded by the European Union

4 Predictive Layer

SafeOPS' Predictive Layer planned and performed all technical tasks related to the **development of an AI/ML solution** for the decision intelligence concept. The following section presents a brief overview of the methodology developed and followed for the development of the ML solutions in SafeOPS. For a more detailed description we advise the reader to refer to deliverable 4.1 (D4.1) [18] and 4.2 (D4.2) [19] of the project.

The tasks of the predictive layer align with two sections of the EASA guidance, the **learning assurance** and the **AI explainability**. In the following we will briefly describe the SafeOPS approach on both topics and compare the work performed with the EASA guidance on learning assurance and AI explainability. Both sections state numerous objectives, some of which reach well beyond the scope of SafeOPS. Therefore, this document focus on the relevant objectives regarding the exploratory research state of SafeOPS. Nevertheless, before starting the initial development phase of an AI constituent, we recommend familiarizing with all objectives, as with increasing targeted TRLs they become relevant as well.

4.1 Learning Assurance

The two relevant factors from the SafeOPS perspective regarding learning assurance are the IT infrastructure and the AI/ML development workflow, defined in the following subsections.

4.1.1 SafeOPS IT Infrastructure

For the necessary IT-infrastructure, SafeOPS relies on DataBeacon. This infrastructure, initially developed in the SafeClouds.eu project [20], handles data ownership, confidentiality, and data protection. The key element of the SafeOPS infrastructure is a data lake repository for the storage of all the relevant data during the different phases of the development. A data lake is a centralised repository that allows us to store structured and unstructured data in its original format and at any scale. Some main advantages of using a data lake over more traditional data storages are the easy adaptability to an increasing volume of data, the design enabling to handle various data sets and formats at the same time and the elimination of problems with data silos (such as data duplication) by giving downstream users a single place to look for all data sources. For the creation of the data lake as well as for the development of the different phases of the project, some of the best practices of the industry were followed, in particular for the management of the data the Bronze/Silver/Gold scheme was used. An overview of this scheme can be seen in Figure 1. The use of this scheme allowed us to efficiently organise the data according to their degree of transformation and ensure the quality and verification of the data for usage in the predictive models. According to this scheme, three main levels are defined depending on the state of the data:

- **Bronze**: This includes all the data in its original format (raw data). No processed data is stored at this level.
- **Silver**: This level includes data that have already undergone some transformation (e.g. filtering, outlier removal, duplication removal, cleaning, etc.) and can be used for data exploration analysis and problem definition.

Co-funded by the European Union

- **Gold**: This level contains the highest quality data for the ML problem being addressed. All the necessary aggregations and feature engineering has been done and it is ready for its use in the predictive models.

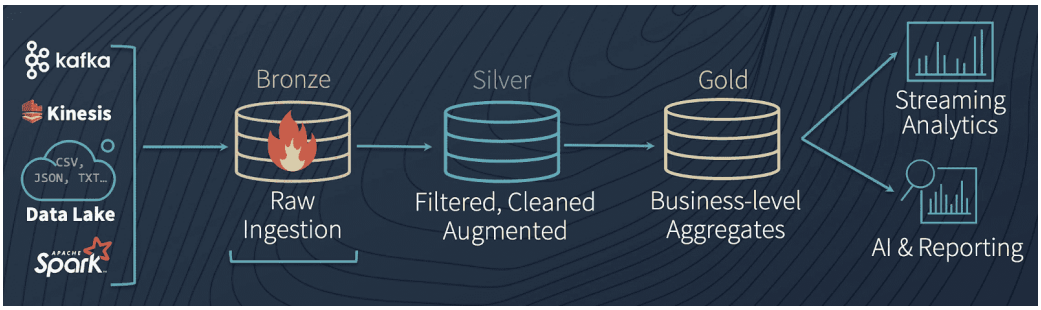

Figure 1 Bronze/Silver/Gold data lake schema [21]

4.1.2 SafeOPS AI/ML Development Workflow

There is no single best workflow, but rather it must be adapted to the needs of the project. However, there are a series of proven best practices that should be followed. The workflow for the predictive layer is as shown in Figure 2. This workflow is inspired by the **Cross-Industry Standard Process for Data Mining (CRISP-DM)** [22]**.** We can distinguish 5 main phases:

1. **Problem definition + Data collection:** The first and most critical step is to define and understand the problem to be solved from an operational perspective. As mentioned previously, this work was carried out in WP2. Once the problem is understood from an operational point of view it has to be translated into a definition from a technical point of view. Therefore, based on the operational layer's work, different possible case studies were defined from an AI/ML perspective based on the expected output, the prediction horizon, geographical horizon, data availability and labelling. With the AI/ML case studies defined and the initial list of data sources identified, data collection can start.

2. **Data processing and preparation + Exploratory data analysis:** The second phase is where the data is transformed from its original format to that needed to train the prediction models. This includes a first step of understanding the data to be worked with. This includes an initial exploration of the data as well as an initial verification phase to check the quality of the data after which we can start with the preparation of the data by the use of Extract, Transform and Load (ETL) pipelines. Once the data has been cleaned (e.g. error correction, outlier detection,...) the final datasets are built by merging the different data sources. At this stage, feature engineering is performed, generating new parameters by combining existing ones. This results in what we call "Clean Data", which are the datasets ready to be used in the training of the prediction models.

3. **Modelling:** Starting from the 'clean' data set, the modelling phase is undertaken. The modelling techniques best suited to the needs of the problem should be selected, and documented together with establishing any possible modelling assumptions. A procedure or mechanism for testing the quality and validity of model results should also be established. Once trained, model performance is evaluated according to the previously established success criteria.

Co-funded by the European Union

4. **Evaluation:** The evaluation in the previous phase focuses more on the technical performance of the model, such as the accuracy and generality of the model. This phase should assess the degree to which the model meets the operational objectives and tries to determine if there are any reasons why the model does not perform as desired. Possible next steps should also be identified, listing the reasons for and against each option.

5. **Deployment:** The last phase of the ML workflow would be the deployment of the solution in a real-world scenario. In the deployment phase, based on the results of the assessment phase, the relevant deployment strategy will be determined. A monitoring and maintenance plan should also be established indicating the necessary steps and the appropriate monitoring and maintenance strategy. This phase was outside the scope of the SafeOPS project.

6. **Iteration:** We can see in the workflow the need for constant iterations through all the different phases of the methodology. It is necessary to constantly test new ideas, establish new hypotheses, extract new features or correct detected errors.

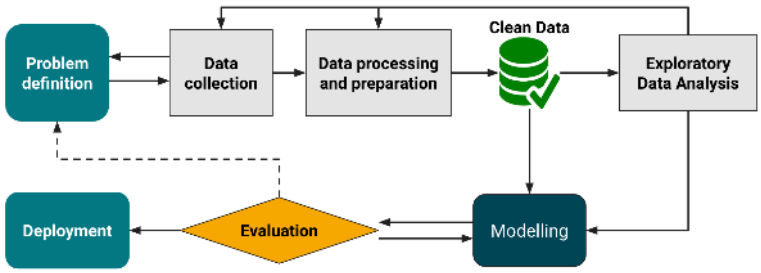

Figure 2 SafeOPS proposed Machine Learning workflow [18]

4.1.3 EASA Learning Assurance Process

The "learning assurance" section provides a framework, which covers "a*ll of those planned and systematic actions used to substantiate, at an adequate level of confidence, that errors in a data-driven learning process have been identified and corrected such that the system satisfies the applicable requirements at a specified level of performance and provides sufficient generalisation and robustness guarantees.*" [11] In addition, EASA proposes a **W-shaped development process**, see Figure 3, adapted to new data-driven learning approaches and building on the more traditional V-shaped development cycle required for the assurance of the development of non-AI/ML components.

As can be seen in Figure 3, the proposed learning assurance framework aims to cover the specific phases of learning processes and takes the highly iterative nature of certain phases of the learning process into account. The workflow in SafeOPS, detailed in section 4.1.2, focused on the initial iterations of the left half of the process, up to first prototypical implementations of a go-around prediction models. The resulting performance metrics were fed back to the operational layer, to validate the initial use cases defined in D2.1 [12].

EUROPEAN PARTNERSHIP

Co-funded by the European Union

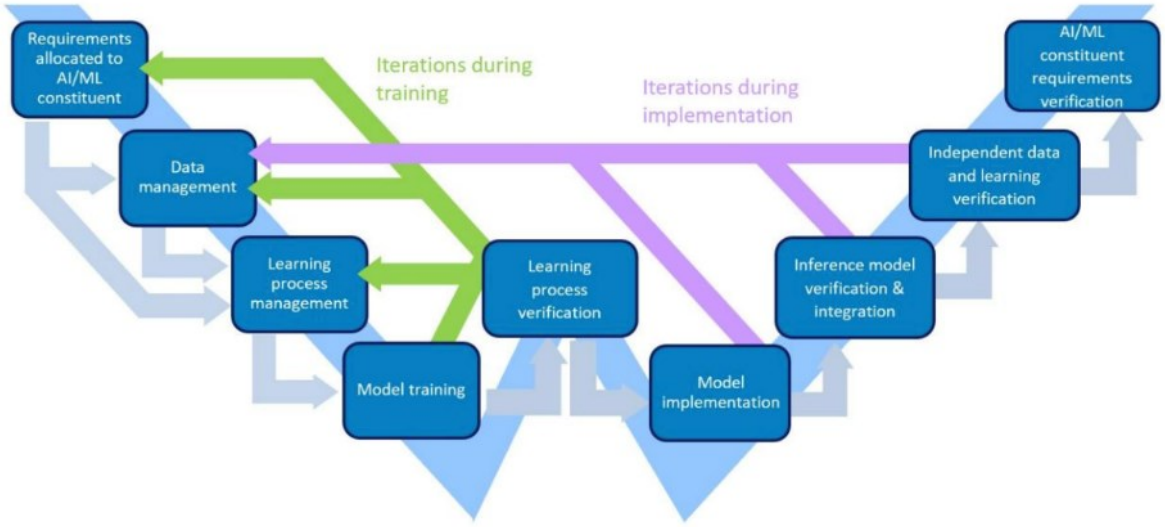

Figure 3: Iterative nature of the learning assurance process [11]

4.1.3.1 Requirements and Architecture Management

For the AI/ML related requirements, EASA states EASA Objective 13 and EASA Objective 14.

EASA Objective 13 [11]

EASA Objective ID	Objective
Objective DA-02	Documents should be prepared to encompass the capture of the following minimum requirements: • safety requirements allocated to the AI/ML constituent • information security requirements allocated to the AI/ML constituent • functional requirements allocated to the AI/ML constituent • operational requirements allocated to the AI/ML constituent, including ODD and AI/ML constituent performance monitoring and data-recording requirements • non-functional requirements allocated to the AI/ML constituent • interface requirements

This objective was partly achieved in SafeOPS, while safety, information security and interface requirements were out of the scope of the current project, functional, operational, and non-functional requirements were developed in D2.1. Also, the DataBeacon infrastructure ensures scalability, which is an important non-functional requirement.

EASA Objective 14 [11]

EASA Objective ID	Objective

Co-funded by the European Union

Objective DA-03	The applicant should describe the system and subsystem architecture, to serve as reference for related safety (support) assessment and learning assurance objectives.

An initial system architecture is described in D2.1, however in case of further work on the concept, a detailed system architecture is one of the most pressing tasks. Especially the interfaces of the AI/ML constituent with the visualization of the information should be described, which would also enable a more in-depth safety assessment, beyond the impact on operational safety, which is discussed in the risk framework in section 3.

4.1.3.2 Data Management

This is the first phase in terms of data lifecycle management. It starts with the definition of specific to data management requirements, in particular **data quality requirements (DQRs)** to be defined for the different types of data selected. These include, among others, to establish the accuracy of the data, the resolution, the confidence that the data has not been corrupted and the traceability of the data.

EASA Objective 15 [11]

EASA Objective ID	Objective
Objective DM-01	The applicant should capture the DQRs for all data pertaining to the data management process, including but not limited to: • **the data needed to support the intended use**, • the ability to determine the origin of the data, • **the requirements related to the annotation process,** • **the format, accuracy and resolution of the data,** • **the traceability of the data from their origin to their final operation through the whole pipeline of operations,** • the mechanisms ensuring that the data will not be corrupted while stored or processed, • **the completeness and representativeness of the data sets,** • **the level of independence between the training, validation and test data sets.**

In SafeOPS, we focused on the bold points of EASA Objective 15, however not in a requirements-based approach. It relates to the second part of the SafeOPS workflow **Data processing and preparation + Exploratory Data Analysis,** described in section 4.1.2.

We **identify this as the most important task**, in case of further work on the concept. Following the "Garbage In, Garbage Out" principle, a thorough documentation of the input data to an AI/ML solution is key for people not directly involved in the AI/ML constituent development, to assess claims and validity of the AI/ML solution. It is therefore paramount to enabling trust in the AI/ML solution for stakeholders, end users, human factors experts, the research community, the funding organisations and most importantly (hopefully at some point) the regulatory bodies. Furthermore, the training data set must demonstrate coverage over the complete operational domain in a sufficient resolution, which is captured in EASA Objective 16.

EASA Objective 16 [11]

EASA Objective ID	Objective
Objective DM-03	The applicant should identify data sources and collect data in accordance with the defined ODD, while ensuring satisfaction of the defined DQRs, in order to drive the selection of the training, validation and test data sets.

SafeOPS has performed a Data Exploration in step 2 of the process defined in section 4.1.2. This marks an initial step toward EASA Objective 16, however, as stated for Objective DM-01, for future work, this needs to be performed in a stricter, requirements-based process which allows verification, also from outside of the development team.

Also important, EASA Objective 17, covers the labelling process.

EASA Objective 17 [11]

EASA Objective ID	Objective
Objective DM-04	Once data sources are collected, the applicant should ensure the high quality of the annotated or labelled data in the data set.

In SafeOPS, we developed a labelling algorithm to divide the pre-processed data set into landing approaches or go-arounds. We visually checked the go-around labelled data in spot tests and compared numbers of go-arounds detected in certain months in our data, with reference data from an ANSP. This twofold approach is important, especially in the case of highly imbalanced data. By comparing the number of labelled go-arounds in our data with a reference source, we could determine that we are not systematically mislabelling actual go-arounds as landings. This is important, as visually inspecting the landing-labels is nearly impossible, given the high imbalance of landings vs. go-arounds. The spot test on the other hand can demonstrate, that we are not systematically mislabel actual landings as go-arounds. However, as for Objectives DM-01/03, future research on this concept should develop more detailed requirements regarding labelling quality, which in turn allow trust-building verification actions in the overall data quality.

Regarding pre-processing and feature computation, EASA Objective 18 summarizes two objectives.

EASA Objective 18 [11]

EASA Objective ID	Objective
Objective DM-06	The applicant should define and document pre-processing operations on the collected data in preparation of the training
Objective DM-07	When applicable, the applicant should define and document the transformations to the pre-processed data from the specified input space into features which are effective for the performance of the selected ML algorithm.

Co-funded by the European Union

Both objectives, summarized in EASA Objective 18, are covered through the **Data processing and preparation + Exploratory Data Analysis,** detailed in section 4.1.2, in the SafeOPS workflow. Additionally, we propose for decision intelligence applications the following SafeOPS Guideline 7.

SafeOPS Guideline 7

SafeOPS Guideline ID	Guideline
SO.GL.7	In case of foreseen real-time predictions, the applicant should capture requirements, specifying the necessary online computation demands of data pre-processing and feature computing tasks.

In SafeOPS, so far we only have performed the pre-processing actions offline. As the concept finally wants to present real time risk information to the end users, an important next step is to demonstrate that pre-processing and feature computing for an approach can be performed online, allowing the envisioned real time predictions. This is therefore an open task for future work.

Common practice in the AI/ML community, however important is EASA Objective 19.

EASA Objective 19 [11]

EASA Objective ID	Objective
Objective DM-09	The applicant should distribute the data into three separate and independent data sets which will meet the specified DQRs: • the training data set and validation data set, used during the model training • the test data set used during the learning process verification, and the inference model verification

EASA Objective 19 is implemented in the **Evaluation** step of the SafeOPS workflow, defined in section 4.1.2.

4.1.4 Learning Process Management

This phase includes the preparatory steps prior to the phase of formal formation of the solution. Among these we can highlight the definition of learning and training requirements in aspects such as model selection, initialisation strategy, hyper-parameters, or cost/loss function selection. The relevant objectives are summarized in EASA Objective 20.

EASA Objective 20 [11]

EASA Objective ID	Objective
Objective LM-01	The applicant should describe the AI/ML constituents and the model architecture.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Objective LM-02	The applicant should capture the requirements pertaining to the learning management and training processes, including but not limited to: • model family and model selection • learning algorithm(s) selection • cost/loss function selection describing the link to the performance and safety metrics • model bias and variance metrics and acceptable levels • training environment (hardware and software) identification • model parameters initialisation strategy • hyper-parameters and parameters identification and setting • expected performance with training, validation and test data sets.
Objective LM-05	The applicant should document the result of the model training
Objective LM-09	The applicant should perform an evaluation of the performance of the trained model based on the test data set and document the result of the model verification.

The objectives summarized in EASA Objective 20 are partly covered in the **Modelling** step of the SafeOPS framework. It is important to state that within SafeOPS we tried various different models for predicting go-arounds and performed a benchmark study to choose the most promising for our endeavour. The process is documented through the IT-Infrastructure of SafeOPS, which also provides versioning of source code and results. At this stage of the project however, only very general requirements for the learning management and training process were defined. With an initial model selected in the scope of this project, refining these requirements would be part of the future work of SafeOPS.

4.1.5 Implementation, Integration and Verification

The listed steps cover the second half of the process defined in Figure 3. They are out of the current SafeOPS scope but are summarized briefly to complete EASA's proposed process. As SafeOPS has not worked on these steps, we will not list the objectives for each proposed step, as we have no experience in implementing them.

Trained model implementation: This marks the first phase corresponding to the implementation of the developed AI/ML solution. This phase aims at transforming the trained model into a model that is executable on the target hardware. This implementation follows different steps in which transformations of the trained model can be performed:

- Model conversion
- Model optimisation
- Model deployment

Inference model verification: This phase has the objective of verifying that the new inference model has an equal/similar performance that the original trained model. Any differences in performance need to be adequately explained. In addition, provision must also be made to verify that the properties of the model have been preserved. Finally, it also includes a first set of typical software verification steps.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Data and learning verification: This phase corresponds to the last step of the data management cycle. The objective is to verify that the datasets used in the development process have been correctly managed. While it should be stressed that this does not imply waiting until the end of the process to initiate this phase, given the highly iterative nature of learning processes, it does indicate that the final verification of the datasets can only take place once the inference model has been successfully verified in the hardware target.

Verification of (sub)system requirements allocated to the AI/ML constituent: This phase covers the verification of the AI/ML component fully integrated into the overall system. It is considered to be covered by traditional assurance methodologies.

Configuration management: During the development process, configuration management principles must be applied at all times to the lifecycle data of the AI/ML components such as versioning, change control, reproducibility or problem reporting.

Quality and process assurance: Quality and process assurance is an integral process in the development process that aims to ensure that the objectives of life cycle processes are met and that activities have been completed as planned (or that deviations have been addressed adequately).

4.2 AI Explainability

In the following, the approach towards AI Explainability of SafeOPS and the EASA is compared. SafeOPS focused on the end-users, whereas the EASA guidance generally addresses

4.2.1 SafeOPS Approach

In SafeOPS, we use the terms interpretability and explainability interchangeably, meaning **"the degree to which a human can understand the cause of a decision/prediction"** and equate both terms in the sense that they both refer to the "ability to provide information about the predictions made by a ML model" [23]. In Deliverable 4.2, an analysis of the different types of explainability techniques currently available in AI/ML was carried out. Focus was laid on the difference between **Model-specific** and **Model-agnostic** methods. Model-specific techniques are those which are specific to a single type of ML model while Model-agnostic techniques are those that can be used on any ML model and are applied after the model has been trained. A distinction can then be made between **Global Model-agnostic** techniques and **Local Model-agnostic** techniques. The Global Model-agnostic techniques focus on the features that have the most impact on all of the model's target outcome. It provides a high-level understanding of how the model works. Local Model-agnostic methods focus on providing specific interpretability on individual predictions of the model. The use of this technique can, in the case of SafeOPS, provide real-time feedback to an ATCO as to why the model is predicting a go-around. In addition, it also allows for a more detailed audit of the model and provides a justification as to why a certain prediction was made. It was therefore decided to use the Local Model-agnostic techniques to explore the explainability of the developed predictive models. The technique known as **SHAP (SHapley Additive exPlanations)** was used. SHAP is an explainability/interpretability technique based on Shapley values developed to provide explanations of individual predictions [24], [25]. It focuses on trying to explain the prediction of a particular instance by calculating the contribution of each feature to the prediction. Even though it is a Local Model-agnostic technique, the explainability results obtained from SHAP allow us to obtain the overall importance (global) of the features by calculating the mean of the absolute Shapley values per feature across all data. In this way, it allows us to obtain both local explanations for particular prediction as well as a high-level analysis of the importance and

Co-funded by the European Union

effect of the different features. Figure 4 shows an example of the global explainability results (left) as well as the results for a specific prediction (right) obtained for Airport 2 at 2NM from the runway threshold.

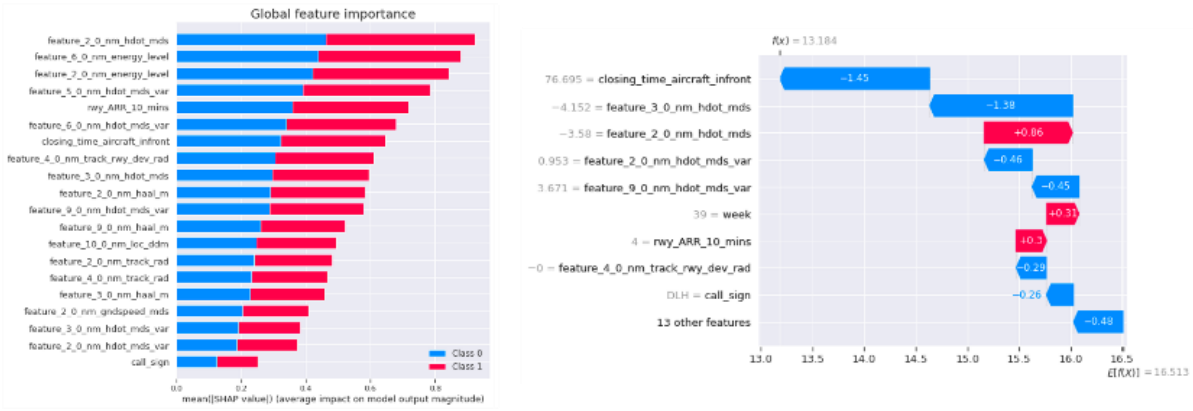

Figure 4 Airport 2 - 2NM - SHAP explainability results (left: Global feature importance; right: local explainability)

4.2.2 EASA Approach

Section 4 of the EASA AI guidelines aims to address the question that arises with the introduction of AI/ML based systems in the decision making process as to how the end users of such tools will understand and interpret the results and reasoning of such systems. The guidelines define explainability as "Capability to provide the human with understandable, reliable, and relevant information with the appropriate level of details and with appropriate timing on how an AI/ML application is coming to its results". In addition, the guidelines also distinguishes between two types of explainability based on the end-user profile and their needs. These are:

- Generate the information required to make a AI/ML model understandable.
- Provide the required information for the user to understand how the systems came to its results.

This classification is mainly based on the condition that different users/stakeholders have different needs and motivations regarding the required level of explainability of an AI/ML solution. It is established that there is an overlap between the motivations of the stakeholders involved in the development as well as in the post-operational phases. In both cases users are more interested in obtaining a very detailed level of explainability about the inner workings of the AI/ML system. In contrast, end-users tend to need more adequate explanations of the operations and results of such AI/ML systems. Figure 5 provides a summary of the needs and motivations of these two defined groups.

Co-funded by the European Union

Development & Post-operation	Operation
▪ Develop system trustworthiness ▪ Establish causal relationships between the input and the output of the model ▪ Catch the boundaries of the model and help in its fixing ▪ Highlight undesirable bias (data sets and model bias) ▪ Allow the relevant receivers identify errors in the model ▪ Enable recording of relevant data to support continuous analysis of the AI-based system behaviour	▪ Contribute to building trust for the end user ▪ Contribute to predicting AI behaviour ▪ Contributing to understanding actions/decisions

Figure 5 Needs and motivations of users for AI explainability [11]

In accordance with the two previously defined groups (Development & Post-Ops and Operation) the EASA AI guidelines provides a series of specific requirements for explainability.

1. **Development & post-ops AI explainability:** Explainability is driven by the needs of stakeholders involved in the development cycle and the post-operational phase:

    a. Target audience: The need for high detail explainability of AI/ML systems concerns a wide range of stakeholders including engineers, certification authorities and safety investigators.

    b. Need for explainability: In addition to the previously mentioned needs/motivations such as learning assurance or trustworthiness analysis, these stakeholders often require a deeper level of knowledge regarding the details of AI/ML system design.

    c. Objectives: The guidelines provide a series of objectives related to the AI/ML explainability such as:

        i. Learning assurance: This is a key prerequisite to ensure confidence in the performance and intended function of AI/ML system. Among other things, for each identified stakeholder, requirements should be set in terms of the desired level of explainability, either for the AI/ML model itself (a priori/global explanation) or for the outcomes of the AI/ML model (post hoc/a posteriori/local explanation).

        ii. Data recording capability: Adequate means must be developed to record the required operational data to be able to provide explanations of the behavior of the AI/ML system in a post-operations scenario. This requires the recording of data for the purpose of monitoring the safety of AI/ML systems (as part of safety management and/or approval of continued operation) as well as the recording of data for the investigation of possible accidents or incidents.

2. **Operational explainability:** In contrast to the previous group, here, explainability is established as the need to provide end users with "understandable" information about how the AI/ML system has arrived at its results.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

a. Target audience: The target audience is considered to be the pilot and co-pilot for air operations, the ATCO and the room supervisor for the ATM domain, and the maintenance engineer for the maintenance domain. These stakeholders are expected to have specific explainability needs in order to be able to use the AI-based system, interact with it and influence its confidence level.

b. Need for explainability: The future trend will be the introduction and use of AI/ML in situations with increasing authority and autonomy. This will inevitably lead to a reduction in the end-user's awareness of the logic behind the automatic decisions or actions taken. This reduced awareness may limit the effectiveness of the interaction and lead to a possible reduction in end-user confidence. To ensure adequate effectiveness of interactions, the AI-based system will need to provide explanations for its automatic decisions and actions.

c. Objectives: The EASA AI guidelines set out a number of objectives as an initial guidance for the design of AI/ML system and its HMI (Human-machine interface). Among other things, it is established that explainability must be provided in a clear and unambiguous manner as well as defining the level of explainability taking into account the characteristics of the task and the situation. In addition, the time at which the explainability will be given to the end user must be defined, taking into account the time criticality of the situation, the end user's need and the operational impact. Finally, AI/ML explainability will also aim to monitor and verify that operating conditions remain within acceptable limits and performance is in line with the expected level.

5 Conclusion

This deliverable is the last technical deliverable of SafeOPS. The goal of this deliverable is, to provide generalized guidelines on the Decision Intelligence concepts for Air Traffic Management. Since Decision Intelligence relies on AI/ML components, we focus especially on the guidance material for their implementation towards certification. During the project's lifetime, EASA published an extensive Guidance Material on AI Applications. Therefore, this deliverable reflected on the processes and workflows, implemented in SafeOPS and compared them to the EASA Guidance on AI Applications. Thereby, we discussed the, for an exploratory research project, applicable/relevant objectives and measured the SafeOPS progress against them. From our lessons learned, which go beyond the EASA guidance, we formulated additional, complementary guidelines.

Overall, we conclude that the SafeOPS approach, which was designed independently from the EASA Guidance, covers most of the relevant objectives at this stage of the process maturity. Especially regarding the **Characterisation of AI Application,** SafeOPS meets the relevant objectives. Furthermore, SafeOPS identified complementary guidelines, originating from the human-integration perspective.

The risk framework, which covers the operational safety assessment and risk considerations of integrating probabilistic information in the Air Traffic Management process also aligns well with the proposed EASA Guidance, regarding **safety assessment**. Regarding **security**, SafeOPS has not foreseen work during this project's life span and therefore, the objectives remain open. A similar status is identified for the **ethics-based assessment**, which should also be considered in case of further investigations of the concept.

Regarding the predictive layer, especially the **learning assurance**, we also observe good alignment with the EASA Guidance. However, for this process, we also identified the most important open tasks, which have to be tackled in future work on the concept. We see the capturing and verification **of Data Quality Requirements** as paramount to enabling trust in the AI/ML solution for stakeholders, end users, human factors experts, the research community, the funding organisations and most importantly (hopefully at some point) the regulatory bodies. We would also recommend that these criteria are integrated in the Maturity Assessment Criteria of the SJU as an addition for AI/ML based projects to ensure compliance from an early development phase.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

6 References

[1] S. Wehmeier und A. S. Hornby, Hrsg., Oxford advanced learner's dictionary of current English, 6. ed., [Nachdr.] Hrsg., Oxford: Oxford Univ. Press, 2004, p. 1539.

[2] A. M. Madni und S. Jackson, „Towards a Conceptual Framework for Resilience Engineering," *IEEE Systems Journal,* Bd. 3, Nr. 2, p. 181–191, 2009.

[3] T. B. Blajev and W. Capt. Curtis, "Go-Around Decision-Makingand Execution Project," 2017. [Online]. Available: https://flightsafety.org/wp-content/uploads/2017/03/Go-around-study_final.pdf. [Accessed 8 9 2019].

[4] T. L. Friedman, „How We Broke the World," New York Times, 31 March 2020. [Online]. Available: https://www.nytimes.com/2020/05/30/opinion/sunday/coronavirus-globalization.html. [Zugriff am 31 October 2022].

[5] B. D. Trump, I. Linkov und W. Hynes, „Combine resilience and efficiency in post-COVID societies," *Nature,* Bd. 588, Nr. 7837, p. 220, 2020.

[6] M. Y. Vardi, „Efficiency vs. resilience," Communications of the ACM, May 2020. [Online]. Available: https://cacm.acm.org/magazines/2020/5/244316-efficiency-vs-resilience/fulltext?mobile=false. [Zugriff am 31 October 2022].

[7] W. A. Galston, „Efficiency Isn't the Only Economic Virtue," Wall Street Journal, 10 March 2020. [Online]. Available: https://www.wsj.com/articles/efficiency-isnt-the-only-economic-virtue-11583873155. [Zugriff am 31 October 2022].

[8] SESAR Joint Undertaking, „European ATM Masterplan," 2020. [Online]. Available: https://www.atmmasterplan.eu/downloads/285. [Zugriff am 30 October 2022].

[9] C. Kozyrkov, "Introduction to Decision Intelligence," 2019. [Online]. Available: https://towardsdatascience.com/introduction-to-decision-intelligence-5d147ddab767. [Accessed 8 9 2019].

[10] E. Humm und C. Abate, „SafeOPS - Human Factors Analyses of the Impact of Providing Probabilistic Risk Information in Real Time," 27 05 2022. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ecb7a660&appId=PPGMS. [Zugriff am 2022 09 30].

[11] EASA, „EASA Concept Paper: First Usable Guidance for Level 1 Machine Learning Applications," 15 2021. [Online]. Available: https://www.easa.europa.eu/en/downloads/134357/en. [Zugriff am 30 09 2022].

[12] L. Beller, M. Pfahler, J. Hartl, P. Kurth, S. Rangger, P. Hernandez und C. Abate, „SafeOPS - Development of Use Cases, User Stories and Requirements," https://cordis.europa.eu/project/id/892919/results, SJU, 2021.

[13] J. Leonhardt, E. Hollnagel, L. Macchi und B. Kirwan, „A White Paper on Resilience Engineering for ATM," Eurocontrol, Brussels, 2009.

[14] S. Shorrock, J. Leonhardt, T. Licu und P. Christoph, „Systems Thinking for Safety: Ten Principles," Eurocontrol, Brussels, 2014.

[15] C. Abate und E. Humm, „SafeOPS - Integrated Risk Framework," 15 07 2022. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5eedd8d00&appId=PPGMS. [Zugriff am 30 09 2022].

[16] B. Rabiller, N. Fota und L. Carbo, „Guidance to Apply SESAR Safety Reference Material," Eurocontrol, 2018.

[17] AI HLEG, „Ethics Guidelines for Trustworthy Ai," [Online]. Available: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419. [Zugriff am 31 October 2022].

[18] P. Hernandez, L. Beller, P. Koppitz und C. Argerich, „SafeOPS - Complete Data Pipeline Description and Machine Learning Solution," 18 05 2022. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ec5dac75&appId=PPGMS. [Zugriff am 30 09 2022].

[19] P. Hernandez, C. Argerich und L. Beller, „SafeOPS - Human Interpretability Framework for the Selected User Stories," 10 08 2022. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5efe602ee&appId=PPGMS. [Zugriff am 30 09 2022].

[20] European Union, „CORDIS EU research results - SafeClouds.eu," 30 09 2019. [Online]. Available: https://cordis.europa.eu/project/id/724100/results. [Zugriff am 20 08 2020].

[21] H. Brenner und D. Lee, „Productionizing Machine Learning with Delta Lake," databricks, 14 August 2019. [Online]. Available: https://www.databricks.com/blog/2019/08/14/productionizing-machine-learning-with-delta-lake.html. [Zugriff am 31 October 2022].

[22] C. Shearer, „The CRISP-DM Model: The New Blueprint for Data Mining.," in *Journal of Data Warehousing*, 2000.

[23] T. Miller, „Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence,* Bd. 267, p. 1–38, 2019.

[24] L. S. Shapley, Notes on the N-Person Game - I, RAND Corporation, 1951.

[25] S. M. Lundberg und S.-I. Lee, „A Unified Approach to Interpreting Model Predictions,“ Bd. 30, Curran Associates, Inc, 2017.

Co-funded by the European Union