Haiku
Human AI teaming Knowledge and Understanding for aviation safety

Deliverable N. 7.3

Validation of the SHS case-based approach in case studies

Authors:

Paola Lanzi (DBL), Nikolas Giampaolo (DBL), Elisa Spiller (DBL)

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Abstract

This deliverable presents the final validation of the SHS-L (Safety, Human Performance, Security, and Liability) framework developed within the HAIKU project, applied across six different Use Cases based on respective Intelligent Assistants (IAs). Building upon the methodology in Deliverable 7.2, this report offers an in-depth assessment of human-AI interactions across varying levels of automation, task complexity, and operational settings analysing the Key Performance Areas (KPAs) identified for each Use Case, considering as well differences in Technology Readiness Levels (TRLs).

The results highlight recurring SHS-L vulnerabilities that are inherently context-dependent, shaped by factors such as the IA's functional role, interface design, human-AI coordination mechanisms, and the adequacy of user training. These findings underscore the importance of tailoring mitigation strategies to specific use case configurations and operational constraints.

This deliverable reaffirms the necessity of addressing SHS-L dimensions in a holistic and anticipatory manner, from the earliest stages of design through to deployment and operational use. The validated HAIKU framework provides a robust, scalable, and human-centred methodology to support the safe, effective, and legally compliant integration of IAs across the aviation ecosystem.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Information table

Deliverable Number	7.3
Deliverable Title	Validation of the SHS case-based approach in case studies
Version	2.0
Status	Final
Responsible Partner	DBL
Contributors	Paola Lanzi, Nikolas Giampaolo, Elisa Spiller
Contractual Date of Delivery	August 31st, 2025
Actual Date of Delivery	September 2nd, 2025
Dissemination Level	Public

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Document history

Version	Date	Status	Author	Description
1.1	14/03/2025	Draft	Nikolas Giampaolo Paola Lanzi Elisa Spiller	Outline
1.2	16/05/2025	Draft	Nikolas Giampaolo Paola Lanzi Elisa Spiller	Draft
1.3	02/07/2025	Draft	Nikolas Giampaolo Paola Lanzi Elisa Spiller	Draft
1.4	8/08/2025	Review	Filippo Tomasello	Review with minor comments
1.5	02/09/2025	Review	Theodore Letouze	Review with minor comments
2.0	03/09/2025	Final	Nikolas Giampaolo	Final version for submission

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

List of acronyms

Acronym	Definition
AI	Artificial Intelligence
ATM	Air Traffic Management
CAT	Commercial Air Transport
CIA	Confidentiality, Integrity, and Availability
CRM	Crew Resource Management
EASA	European Union Aviation Safety Agency
EU	European Union
HAZOP	Hazard and Operability Study
HP	Human Performance
HMI	Human-Machine Interface
IA	Intelligent Assistant
KPA	Key Performance Area
OSD	Operational Sequence Diagram
SA	Situational Awareness
SecRAM	Security Risk Assessment Methodology
SHS-L	Safety, Human Factors, Security, Liability
TRL	Technology Readiness Level
UC	Use Case
UAM	Urban Air Mobility

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Table of contents

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

1. Introduction

1.1. Scope of the document

This deliverable presents the final outcomes of Task 7.3 “Safety, Security and Human Factors Analysis” of the HAIKU project. Building upon the methodology defined in D7.2, the analysis adopts a multidisciplinary approach to evaluate the interaction between human operators and AI systems in operationally relevant scenarios. The objective is twofold: (1) to validate the effectiveness and applicability of the SHS-L (Safety, Human Performance, Security, Liability) assessment framework, and (2) to identify critical vulnerabilities, operational and legal risks, and human-system interaction issues for six different Use Cases (UCs).

The document presents the results of scenario-based evaluations conducted in collaboration with UC partners, including a detailed categorisation of SHS-L issues and contributing factors. This deliverable builds upon the individual assessment reports listed in Annex A, which include a categorisation of SHS-L issues. The final considerations of the aggregate analysis are presented herein, offering a cross-cutting view of the main vulnerabilities and patterns emerging across use cases.

1.2. Structure of the document

This deliverable is organised into five main sections, each contributing to the finalisation of the HAIKU assessment methodology and the consolidation of project results.

- Section 1 introduces the document, outlining its scope, objectives, and structural organisation.
- Section 2 presents the updated and final version of the HAIKU validation framework.
- Section 3 summarises the assessment results for each of the six HAIKU UCs.
- Section 4 contains a comparative analysis of the six UCs.
- Section 5 concludes the document by presenting a synthesis of the main findings and providing a set of cross-case recommendations.

This project has received funding by the European Union’s Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

2. HAIKU Updated validation framework

The HAIKU validation framework provides a structured and multi-layered approach for assessing AI-enabled Intelligent Assistants (IAs) in aviation. The framework is designed to support the systematic evaluation of AI solutions at various Technology Readiness Levels (TRLs) while promoting compliance with safety, security, human factors, and liability considerations. Unlike traditional validation methodologies that focus solely on technical feasibility, the HAIKU framework embraces a human-centric approach, integrating ethical, societal, and operational dimensions. This validation framework is iterative, allowing for continuous refinement as AI-enabled IAs advance through development stages. By encompassing a broad range of Key Performance Areas (KPAs), the framework ensures that AI systems remain trustworthy, explainable, and aligned with regulatory requirements and relevant consensus-based industry standards.

The validation framework builds upon four interdependent KPAs, ensuring that AI-enabled IAs are assessed holistically:

- **Safety** is assessed to ensure that AI solutions operate predictably and fail safely in high-risk environments, mitigating risks linked to automation complacency and decision errors.
- **Human Performance** is evaluated to ensure AI systems support operators effectively, without increasing cognitive workload, reducing usability, or hindering decision-making processes.
- **Security** considerations focus on ensuring AI systems remain resilient against cyber threats, data breaches, and adversarial attacks.
- **Liability** risk is addressed by aligning AI system and concept design decisions with the existing aviation regulatory ecosystem, including EASA guidance on AI, the EU AI Act, and international aviation law, whose implications on the life-cycle of IAs have been presented in Deliverable D7.2.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

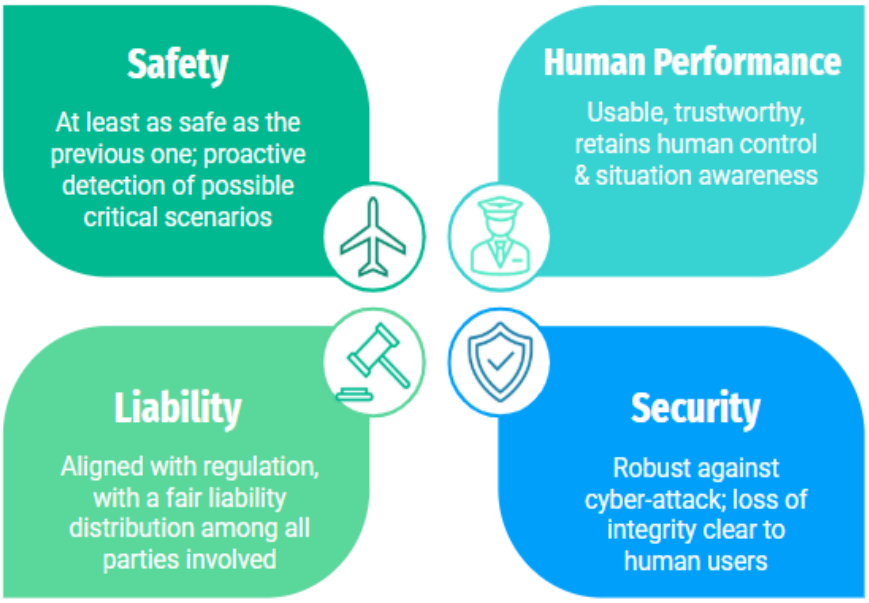

Figure 1. KPAs Dimensions of the Validation Framework

A fundamental feature of the updated validation framework is its scalability across different TRL stages, ensuring incremental validation of AI-enabled IAs. The validation methodology follows a four-stage assessment process to ensure AI-enabled IAs are systematically evaluated. The process begins with **Operational Sequence Diagrams (OSDs),** which map human-AI interactions, identifying cognitive workload imbalances, operational bottlenecks, and decision-making challenges. The second stage applies the **Hazard and Operability (HAZOP) methodology** to systematically assess potential safety risks, system vulnerabilities, and unintended AI behaviours. The third stage involves a comprehensive **risk assessment** across all Key Performance Areas, ensuring AI-enabled IAs meet safety, security, human performance, and liability standards. The final stage focuses on implementing **mitigation** strategies, integrating adaptive safety measures, security safeguards, and trust calibration techniques to ensure AI systems evolve in a human-centred and legally compliant manner.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

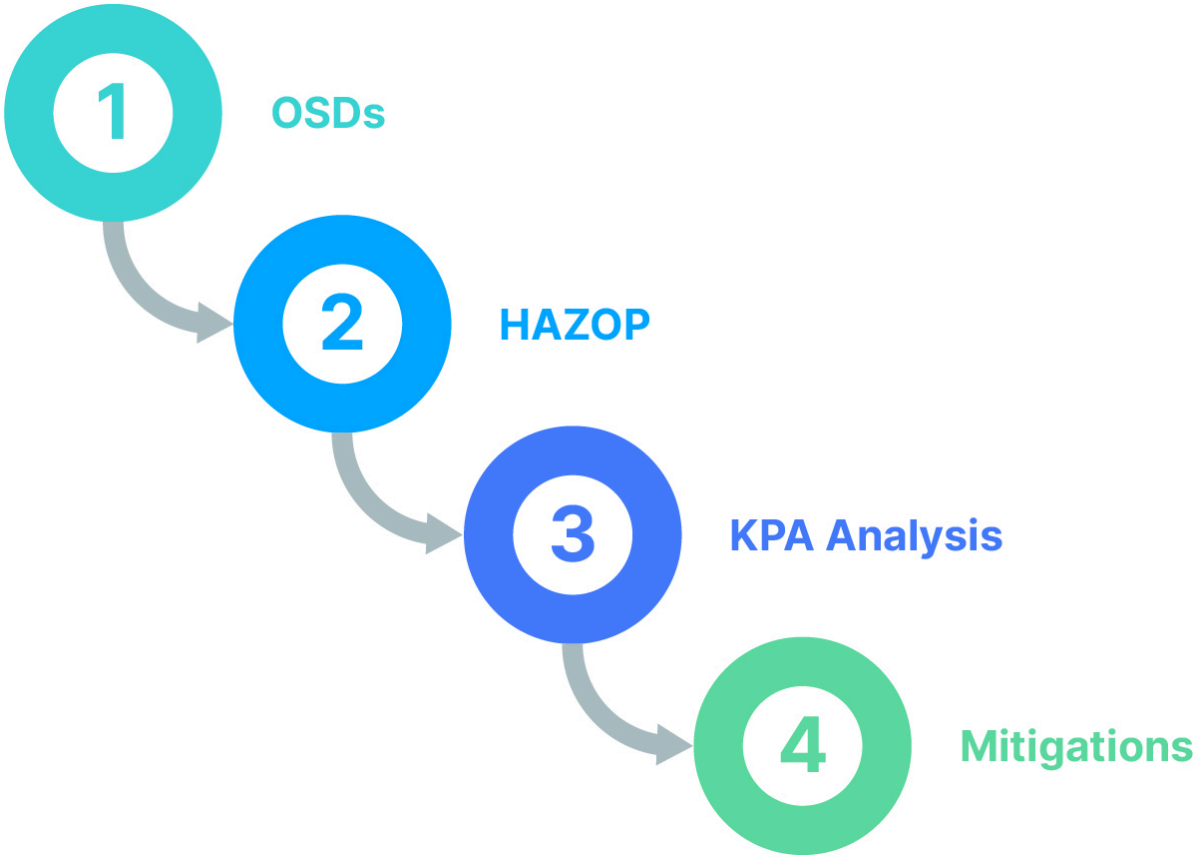

Figure 2. Methodological Steps

A core component of the HAIKU framework is the identification, through the HAZOP analysis, of safety vulnerabilities associated with human-AI interaction, system reliability, and decision-making risks. Safety issues are categorised into three primary types: **lack of preparedness**, **unpredictable or inconsistent decisions**, and **overreliance** on AI outputs. These risks are particularly critical in aviation, where even minor errors can have cascading effects.

These safety vulnerabilities are closely linked to Human Performance (HP) challenges, including interface design flaws, poor communication, misaligned situational awareness, trust deficits, and insufficient training. Such factors undermine operators' ability to interpret, trust, and act upon AI system outputs, especially under time pressure or in complex scenarios.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

The framework also integrates a **liability assessment** based on the Legal Case methodology, addressing product, organisational, and personal responsibility. This ensures that legal risks are considered from the earliest design stages, aligned with AI governance principles and evolving legal framework.

From the **security** perspective, the framework applies the **SecRAM 2.0** methodology to selected use cases (UC1, UC5, UC6), focusing on risks to data **Confidentiality, Integrity, and Availability (CIA)**. Although full threat modelling was constrained by system maturity, the structured evaluation provided valuable insights into potential vulnerabilities.

Finally, the framework defines a set of **mitigation measures** tailored to system stakeholders. These measures aim to promote cross-cutting improvements in data integrity, system reliability, and human-AI collaboration, and are designed to evolve with technological and regulatory maturity.

For a detailed description of methods, assessment categories, and analytical rationale, **please refer to Deliverable D7.2**.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

3. Summary of Use Cases and Assessment Results

This section presents an overview of the six AI-enabled systems (UC1–UC6) assessed within the HAIKU validation framework. Each UC was evaluated using a multi-layered methodology encompassing safety, human performance, liability, and—where applicable—security risk assessments (conducted only for UC1, UC5 and UC6), as well as tailored mitigation strategies. The systems span a broad range of operational domains, from AI-assisted pilot support during startle events (UC1) and dynamic in-flight route planning (UC2), to urban air mobility coordination (UC3), air traffic sequencing in tower environments (UC4), airport-level safety risk analysis (UC5), and public navigation in health-sensitive terminal spaces (UC6). While unified under a shared SHS-L framework, each system is shaped by distinct operational contexts and AI maturity levels. The following summaries describe the core functions of each UC, the specific risks identified through the HAZOP and OSDs across the relevant KPAs, and the mitigation strategies proposed to support safe, explainable, secure, and legally accountable deployment.

UC1 - IA for the flight deck startle response (FOCUS)

Startle reactions occur when a pilot encounters an unanticipated event such as a rapid automation failure, a sudden autopilot disengagement, uncommanded aircraft behaviour, or external hazards like a bird strike, severe turbulence, or an engine failure. Physiologically, these reactions can result in rapid heart rate escalation, increased pupil dilation, breathing irregularities, and erratic gaze patterns, all of which significantly deteriorate cognitive performance and affect control of muscle movements. If not managed effectively, these physiological and psychological responses can delay or impair decision-making, potentially leading to cascading errors that compromise flight safety.

The Startle Response Assistant (FOCUS) functions as a real-time monitoring and intervention system, using flight data, physiological data, and behavioural data to detect when a pilot is experiencing a startle or surprise reaction. FOCUS employs a combination of biometric sensors to track heart rate variability, respiration rate, pupil diameter, and gaze behaviour in conjunction with operational inputs from flight deck systems to determine whether an intervention is required. The IA then provides adaptive support mechanisms, including visual and auditory prompts, procedural recommendations, stress regulation techniques, and cognitive reinforcement

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

strategies. The assistant's role extends beyond simple advice. In critical situations, it serves as a co-pilot-level team member, capable of executing monitoring tasks under human supervision, cross-checking pilot responses, and ensuring procedural compliance. The system must be designed to continuously adjust its intervention strategy based on the evolving state of the pilot, adapting in real time to avoid adding unnecessary cognitive load or introducing distractions.

UC1 - Safety Assessment

An extensive safety-critical event analysis (see Report UC1, Annex A) conducted within the scope of UC1 has identified three primary areas of concern: Lack of Preparedness, Unpredictable/Inconsistent Decisions, and Overreliance on AI Assistance. Across the 23 safety-critical events analysed, 27.3% were attributed to Lack of Preparedness, 72.7% to Unpredictable/Inconsistent Decisions, and 9.1% to Overreliance on AI Recommendations. These issues highlight the complex interplay between human cognitive performance, IA intervention accuracy, and operational safety in the cockpit.

The Lack of Preparedness category encompasses situations in which FOCUS could fail to anticipate or appropriately respond to the pilot's startle reaction, leaving the crew unassisted in critical moments. This could occur due to delayed detection of physiological markers, misinterpretation of pilot state, or a lack of proactive engagement from the pilot in selecting or deactivating the assistant's recommendations. Furthermore, the lack of preparedness extends to the pilot's ability to recognize their own cognitive degradation and appropriately engage with the system when necessary. This underscores the necessity for robust training programs that familiarise pilots with the system's capabilities, limitations, and intervention protocols, ensuring that both the AI and human operators are prepared for unexpected in-flight contingencies.

The Unpredictable/Inconsistent Decisions category presents the greatest risk, as it represents situations in which FOCUS may provide erratic, conflicting, or delayed responses due to software malfunctions, poor interface design, or misaligned risk prioritisation. Instances of IA misjudging the severity of a startle event, failing to recommend the appropriate intervention, or providing contradictory procedural prompts have been identified as major risk factors. The presence of calibration errors, misaligned system thresholds, and unstable feedback loops may lead to incorrect procedural activations or failure to engage critical support mechanisms at the right

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

time. A significant challenge is ensuring that FOCUS consistently aligns with the pilot's mental model and operational context, preventing confusion or cognitive overload that could further degrade flight crew performance.

The primary risk associated with overreliance is that pilots may begin to trust responses indiscriminately, failing to apply independent critical thinking or override the system when necessary. This automation complacency could result in inappropriate procedural execution, misjudgment of the aircraft's operational state, and decreased pilot confidence in their own ability to manage high-stress events. This issue is exacerbated when the system provides incorrect estimates or fails to highlight critical areas of concern, leading to a false sense of confidence in automation-generated recommendations. Addressing this challenge requires refining IA reliability, ensuring system transparency, and implementing operational safeguards that encourage manual intervention when necessary.

UC1 - HP Assessment

The HP analysis within UC1 underscores the intricate relationship between pilots and automated systems, revealing that trust-related concerns constitute the dominant issue, accounting for 70.8% of all HP failures. This is followed by training deficiencies (37.5%), communication breakdowns (12.5%), interface usability challenges (25%), and shared situational awareness deficits (8.3%). Each of these concerns presents significant implications for operational safety, necessitating robust mitigation strategies in the system's design, deployment, and pilot training programs.

The loss of trust in the IA is a direct result of inconsistent AI decision-making, lack of transparency in system logic, and misalignment with pilot expectations. Pilots develop mental models based on their experience, training, and operational context and when recommendations fail to align with these models, pilots may reject system guidance entirely, even in situations where it could be beneficial. For example, in Unpredictable/Inconsistent Decision U11 (see Report UC1), the AI may fail to correctly interpret gaze-tracking data, leading to incorrect visual cues that mislead the pilot. This, in turn, could erode confidence in FOCUS's ability to assist, contributing to cognitive overload and increased manual workload, especially in high-stress scenarios. Another major trust-related concern arises in Unpredictable/Inconsistent Decision, where poorly calibrated biometric readings result in the IA misidentifying a startle response, leading to inappropriate or untimely interventions. This creates a scenario where the pilot may no longer feel the IA is reliable, reducing overall system

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

effectiveness. Trust issues are particularly dangerous in highly automated environments, as pilots who lose confidence in guidance may hesitate to engage with the system at all, defeating the purpose of the assistant's intervention capabilities.

Training deficiencies contribute to HP failures, emphasizing the need for structured and recurrent training programs that ensure pilots fully understand the IA's functionalities, limitations, and appropriate use cases. This highlights a gap in operational readiness, where pilots either fail to interpret system outputs correctly or are unable to manually override recommendations in a timely manner.

The communication breakdowns further amplify these challenges. If the IA fails to provide correct breathing guidance due to software malfunctions or misinterpretation of biometric data, FOCUS can lead to an increased stress response rather than mitigation. This scenario deserves particular attention, as the pilot expects support that is not delivered effectively, increasing frustration, cognitive overload, and the likelihood of performance errors.

Interface usability challenges indicate that the Human-Machine Interface (HMI) must be designed to ensure intuitive engagement and accurate visual representations of recommendations. The IA may provide incorrect visual cues due to calibration errors, causing the pilot to misinterpret critical information and respond incorrectly. Poorly designed interfaces that do not align with natural pilot workflows create additional distractions, increase decision-making delays, and reduce overall system effectiveness.

Finally, shared situational awareness remains a persistent challenge, particularly in cases where the IA's interpretation of flight conditions diverges from the pilot's assessment. FOCUS may provide incorrect estimates, which the pilot blindly follows without cross-checking other data sources. This demonstrates the danger of automation complacency, where pilots begin to defer too much authority to outputs, leading to misaligned decision-making and potential safety risks.

UC1 - Security Assessment

The security assessment for UC1 was carried out using the SecRAM 2.0 methodology, adapted from SESAR, and applied to the startle response assistant. Although the technological maturity of the system limited the application of all SecRAM phases, the evaluation successfully completed the initial steps of asset identification and impact assessment. The core objective was to understand how malicious or accidental threats

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

could compromise system confidentiality, integrity, and availability (CIA), and what consequences such degradation would have on operations and safety.

The assessment began by identifying the primary assets — namely, the AI-based startle detection and assistance module, physiological sensors (e.g., heart rate monitors, pupil dilation tracking), the Human-Machine Interface (HMI), and the AI's decision-support algorithms. These were complemented by supporting assets, including the data transmission infrastructure, cockpit integration layer, and data repositories used for model training and validation.

From there, a CIA-based impact assessment was conducted. In terms of confidentiality, a breach could result in unauthorised access to sensitive biometric data collected from pilots — data which, by its nature, is highly personal and subject to both aviation-specific data governance and broader EU data protection regulations (e.g. GDPR). An attack targeting integrity was deemed particularly consequential, as manipulation of sensor data or the decision engine could lead to false detection of startle conditions or inappropriate activation/deactivation of support procedures. For instance, a false positive might trigger unnecessary system responses, potentially increasing pilot confusion during flight. A false negative could result in the assistant failing to activate in a critical moment. The availability of the assistant was likewise assessed as vital. Denial-of-service (DoS) scenarios — whether via signal jamming, flooding, or internal resource exhaustion — could block the IA's real-time activation, thereby nullifying its support capabilities during high-stress flight phases.

Impacts were evaluated using the standard SecRAM impact table, which considers effects on personnel safety, system capacity, performance, economic costs, regulatory compliance, and brand reputation. For UC1, the most critical impacts were those related to personnel and performance, given that the IA is expected to intervene precisely when pilots are in a cognitively degraded state. Any security breach that compromises the assistant's functionality at these moments could directly result in operational failure and jeopardise flight safety.

Due to the current TRL of the system, the assessment did not progress to the identification of specific threat agents or likelihood estimations. However, the preliminary analysis already highlighted the necessity of embedding strong cybersecurity measures from the earliest stages of system development as mandated by EU/EASA Part-IS.

UC1- Liability Analysis

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

The liability considerations surrounding UC1 are complex and multifaceted, involving multiple stakeholders, including technology providers, aircraft manufacturers, airline operators, regulatory bodies, and pilots. Given that AI-based systems introduce a layer of automated decision-making into human-controlled environments, determining who is accountable for system errors, misinterpretations, or failures becomes increasingly difficult.

At the current level of design maturity, liability primarily falls on the technology provider, as they are responsible for ensuring that the AI system is designed, tested, and validated in accordance with aviation safety standards and the new specific requirements for AI. However, as the system progresses toward full-scale deployment, liability considerations will extend to aircraft operators and individual pilots, necessitating clear regulatory frameworks defining oversight responsibilities, procedural compliance, and acceptable AI decision-making thresholds.

Considering product liability issues, a key liability risk involves design defects and transparency failures, where the AI system fails to accurately communicate its limitations to pilots. According to the guidance provided by the EU AI Act and EASA for the development and deployment of AI in aviation, these systems should be generally classified as high-risk. As a consequence, they shall be designed the enable effective human oversight on systems and operations, allowing end-users to:

- Fully understand the system's capabilities and limitations, including its accuracy thresholds, expected behavior, and potential failure modes.
- Monitor AI operation in real time and detect anomalies, dysfunctions, or unintended performance deviations.
- Avoid over-reliance on outputs, particularly in cases where automation bias could compromise independent decision-making.
- Correctly interpret AI recommendations, considering system design logic, available interpretation tools, and real-world flight context.
- Intervene manually to override or halt AI-driven decisions, ensuring that the final authority in critical scenarios remains with the human operator.

Failure to adequately address these oversight requirements introduces legal risks related to product liability, organizational fault, and pilot negligence in AI-assisted decision-making errors.

Product liability concerns may also emerge in cases where the system fails to perform as intended, resulting in erroneous recommendations, inconsistent interventions, or failure to activate when needed. In Unpredictable/Inconsistent Decision U3, the AI

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

system experiences software malfunctions that delay critical alerts, leading to potentially unsafe conditions for the flight crew. Under the Product Liability Product Directive (PLD) EU 2024/2853, technology providers must prove that all reasonable safety measures were taken during the system's development, validation, and deployment. If such measures were inadequate, liability for resulting accidents may be attributed to the AI providers.

Organisational liability applies to deploying organisations (e.g. aircraft operators) responsible for system operational use, maintenance, and personnel training. If a pilot misuses or fails to correctly engage with the AI system, questions arise regarding whether adequate training was provided. In Lack of Preparedness L5, for example, pilots fail to understand the AI's procedural hierarchy, leading to confusion in manual versus AI-assisted intervention. If training records indicate that the aircraft operator did not provide adequate instruction, the company could be held liable for improper implementation of the system.

Finally, pilot liability is a growing concern in AI-assisted decision-making environments. Under current aviation regulations, the pilot-in-command (PIC) retains ultimate authority and responsibility for flight safety. However, with increased automation, pilots may be exposed to legal risks if they blindly follow the recommendations without cross-verification. In Overreliance O2, the pilot accepts an AI-provided decision without manually verifying flight conditions, leading to an incorrect response. In such cases, liability could shift to the pilot if it is determined that they failed to exercise due diligence in evaluating outputs.

UC1 - Mitigations

The mitigation strategy outlined for UC1 reflects a comprehensive, multi-stakeholder approach tailored to the system's current maturity level. These measures are not confined to individual KPAs, but rather address the overarching challenges identified across safety, human performance, security, and liability analyses. Their aim is to guide the continued development of the startle response assistant in a manner that ensures operational resilience, human oversight, and regulatory compliance.

At the design level, the recommendations for developers and producers focus on ensuring that the assistant's core function — accurate detection of startle reactions in pilots — remains consistently reliable under varied operational conditions. This requires the implementation of safeguards against system degradation and data integrity breaches, including specific protections against data poisoning, adversarial

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

inputs, and confidentiality attacks. These elements are essential not only for securing the assistant's computational reliability but also for ensuring that its performance is trustworthy in scenarios of acute pilot vulnerability. To support this, long-term robustness and cybersecurity must be preserved through design transparency, including the declaration of accuracy metrics and system limitations, and through structured logging of physiological and system data for audit and diagnostic purposes.

Transparency and explainability are emphasised as foundational pillars of user trust and operational usability. Developers are expected to equip the system with interfaces and human-machine interaction modalities that make the assistant's outputs intelligible and actionable under time pressure. This includes enabling pilots to verify, reject, or override AI outputs, and ensuring they are provided with visual or textual explanations of why particular recommendations were made. These interaction pathways are not merely ergonomic considerations — they form part of the system's safety architecture by preserving the pilot's final authority and preventing overreliance or blind compliance. The assistant must be designed to allow pilots to monitor anomalies in real-time, assess the system's logic, and maintain situational awareness even in degraded cognitive states. In this sense, control and interpretability are safety features, not just usability concerns.

On the organisational side, aircraft operators bear responsibility for embedding the assistant into operational routines without compromising existing safety practices. The recommendations emphasise the need for domain-specific training that accounts for variations in crew experience, technical familiarity, and cognitive readiness. Training is not limited to familiarisation with interface commands, but extends to scenario-based exercises that prepare pilots to interpret startle detection alerts and execute appropriate responses, including situations where they must override or disengage the system, ensuring comprehensive pilot training on AI-human interaction principles, refining system communication protocols, and embedding fail-safe mechanisms that allow for human oversight at all levels of AI intervention. Furthermore, structured reporting and feedback mechanisms must be implemented, enabling anomalies or malfunctions encountered in-flight to be documented and escalated. This is particularly critical in early deployments, where iterative refinement based on real-world usage will determine the system's long-term reliability and acceptance. In this context the guidance provided by ICAO for Evidence-Based Training (EBT) in Doc 9995 may be useful.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

UC2 - Flight Deck Route Planning and Replanning (OLIVIA)

The Intelligent Assistant for Flight Deck Route Planning and Replanning (OLIVIA) is developed as a decision-support system designed to assist flight crews in dynamically adjusting flight routes based on real-time meteorological conditions, air traffic constraints, and airport status updates. The IA acts as a cognitive augmentation tool, providing automated hazard analysis, trajectory optimization, and predictive rerouting recommendations that enable pilots to make more informed, data-driven decisions in time-sensitive scenarios.

The IA does not assume authority over flight operations. Instead, it serves as an advisory and supervisory tool, designed to function in two distinct but complementary phases. In the advisory phase, the IA continuously scans environmental and operational parameters, identifying potential hazards such as turbulence, convective weather, or airspace congestion. Upon detecting a threat, it analyses its severity, evaluates possible rerouting strategies, and presents recommendations to the flight crew for validation. In the supervisory phase, OLIVIA actively monitors the pilot's decision-making process, cross-checking selected flight path adjustments against available risk assessments and operational priorities. If a deviation from an optimal trajectory is identified—whether due to an oversight, misjudgement, or failure to account for an emerging hazard—the IA provides contextualized corrective prompts, ensuring that safety margins and regulatory constraints are maintained.

To perform these functions, OLIVIA integrates multiple data streams and computational methodologies that collectively enhance the accuracy and timeliness of flight planning decisions. The system employs digital hazard detection tools, leveraging airborne weather radar, System-Wide Information Management, and Air Traffic Management (ATM) data networks to continuously assess evolving flight conditions. Risk assessment algorithms quantify the potential impact of meteorological threats, such as icing, turbulence, or reduced visibility, classifying them based on severity and likelihood. Using these insights, the IA generates ranked alternative flight paths, balancing safety considerations with fuel efficiency and ATC-imposed constraints. The system also facilitates real-time coordination with ATC and AOCC, ensuring that all revised flight plans comply with regulatory requirements and airspace management priorities. Finally, bidirectional communication protocols are embedded within the system to promote seamless interaction between pilots and recommendations, allowing for manual adjustments, overrides, and iterative refinements in response to evolving conditions.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

UC2 - Safety Assessment

A comprehensive risk evaluation of UC2 identified twelve primary failure scenarios (see Report UC2, Annex A), which were classified into three dominant safety risk categories:

Representing the most prevalent issue in UC2, Unpredictable or Inconsistent AI Decisions account for 63.6% of all safety-critical events. This category encompasses a wide range of system failures in hazard detection, risk prioritisation, and explainability of recommendations. In U1 (see Report UC2 listed in Annex A for further details), OLIVIA significantly underestimated meteorological hazards, compromising safe routing. Events U2 and U3 reflect flawed risk prioritisation, wherein the IA promoted route suggestions that, while perhaps optimal under one parameter (e.g., efficiency), neglected other critical factors (e.g., safety or crew condition), thus leading to unsafe or inefficient outcomes. In U4 and U5, the assistant failed to generate actionable alternatives, increasing pilot workload and complicating time-sensitive decisions. In U6, inappropriate internal weighting mechanisms (e.g., privileging fuel economy over operational safety) further misaligned system outputs with pilot expectations. Most notably, U7 exposed a structural limitation: the absence of operational-level explainability. Without timely and comprehensible justification for its suggestions, OLIVIA's outputs were either disregarded or executed with delay, as pilots lacked the confidence necessary to act upon them. These episodes emphasise the paramount requirement for robust explainability mechanisms and context-sensitive prioritisation strategies within IA systems, particularly those operating in cognitively demanding environments.

The Lack of Preparedness category accounts for approximately 27.3% of the safety-critical events analysed in UC2. These scenarios reveal instances in which pilots might fail to adequately interpret or validate the routing suggestions provided by the Intelligent Assistant (OLIVIA), primarily due to limited training, misalignment with operational priorities, or cognitive overload during demanding flight phases. In event L1, pilots could misjudge meteorological threats such as turbulence or icing, often exacerbated by difficulties in deciphering the IA's outputs, leading to suboptimal assessments of environmental hazards. In L2, misinterpretation of route options resulted in the selection of inefficient or operationally unsuitable paths. Event L3 highlights instances in which the pilot's operational intent—such as avoiding turbulence—could be at odds with OLIVIA's optimisation criteria, such as fuel savings,

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

thereby introducing inefficiencies and eroding trust. In L4, AI recommendations were dismissed entirely, typically due to an inadequate understanding of the assistant's rationale or persistent scepticism towards its reliability. These findings underscore the critical need for harmonised decision-making logic between the human operator and the AI, supported by targeted training that cultivates appropriate engagement strategies under various cognitive and operational conditions.

Although less frequent, Overreliance on the Intelligent Assistant remains a consequential safety concern, present in 9.1% of the scenarios. The relevant event (O1) illustrates a behavioural pattern of automation complacency, in which pilots may adhere to OLIVIA's recommendations without performing critical verification. This uncritical acceptance is especially hazardous in contexts where system outputs, though apparently plausible, are based on flawed or incomplete data. The failure to cross-check IA guidance may lead to suboptimal or unsafe route selections and reflects a broader disengagement from active situational assessment. This behavioural drift reinforces the need for training strategies that sustain pilot vigilance, encourage routine verification of AI outputs, and cultivate a balanced human-AI interaction wherein the assistant supports —rather than supplants— human judgement.

UC2 - HP Assessment

The integration of an IA for route planning and replanning within the cockpit introduces significant HP considerations, affecting pilot cognitive workload, situational awareness, trust calibration, and communication effectiveness. The HP analysis of UC2 highlights five key performance degradation domains: trust issues, training deficiencies, interface and interaction design limitations, shared situational awareness discrepancies, and communication breakdowns. Each of these areas directly impacts how pilots engage with the AI system, interpret its recommendations, and execute decisions based on its outputs.

Trust-related concerns represent the most prevalent human performance issue in UC2, accounting for 63.6% of HP-related failures. These concerns primarily stem from inconsistent IA behaviour, unreliable hazard detection, and inadequate operational explainability (OpXAI). When the IA provides conflicting recommendations or fails to justify its suggested flight path adjustments, pilots become increasingly sceptical about the system’s reliability. In scenarios where OLIVIA misidentifies or fails to detect a severe meteorological hazard, pilots may question its decision-making capacity,

This project has received funding by the European Union’s Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

ultimately hesitating to act on route adjustments even when the recommendations are appropriate. This loss of trust creates a cognitive burden on flight crews, forcing them to expend additional mental effort on validating outputs rather than benefiting from its intended decision-support role. In instances where pilots experience multiple inconsistencies over time, they may default to ignoring the system entirely, effectively rendering the AI unusable in critical moments when it could provide the most value.

Training deficiencies are another major contributor to human performance issues in UC2, making up 18.2% of identified failures. Pilots require extensive training to properly interpret and evaluate recommendations, yet current training paradigms in aviation do not always emphasize the cognitive collaboration required for human-AI teaming. In scenarios where the IA suggests an alternative route due to an upcoming storm system, a pilot who has not been adequately trained in how the AI prioritizes hazards and operational constraints may misinterpret the severity ranking of the system's recommendations. This misunderstanding can lead to two distinct outcomes: either the pilot overrides a valid suggestion, resulting in an unnecessary deviation from an optimal route, or the pilot blindly follows the AI's recommendation without performing cross-checks, increasing the likelihood of accepting a suboptimal flight path.

Shared situational awareness discrepancies further exacerbate decision-making challenges, with 9.1% of human performance failures attributed to misalignments between pilot perception and risk assessments. Situational awareness is a multidimensional cognitive process, requiring the integration of visual, auditory, and computational inputs to develop an accurate mental model of the aircraft's operational state. When the IA assesses meteorological hazards and generates route alternatives, it relies on a data-driven analysis of environmental variables, whereas pilots depend on their real-time sensory input, procedural training, and operational experience. If the AI ranks a particular storm system as low-risk based on statistical probabilities while the pilot visually identifies severe convective activity ahead, the mismatch between human intuition and IA-calculated risk assessments can lead to delayed or incorrect decision-making. This is especially problematic when pilots are operating in high-stress, high-tempo scenarios, where any additional uncertainty increases reaction time and cognitive load.

Interface and interaction design limitations introduce further human performance concerns, particularly in how the IA communicates its risk assessments and route recommendations. The IA's HMI must be designed for clarity, intuitiveness, and immediate comprehension, yet multiple scenarios in UC2 have highlighted deficiencies

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

in how information is presented, prioritized, and processed by pilots. For instance, in cases where the IA fails to properly highlight its reasoning behind a suggested route deviation, pilots may dismiss the recommendation due to a lack of perceived urgency. Similarly, if the IA provides an excessive amount of complex data without a clear ranking of priorities, pilots may struggle to distinguish between critical hazards requiring immediate attention and less urgent operational advisories. Poorly structured HMI interactions lead to delayed responses, increased cognitive strain, and greater reliance on secondary validation sources, reducing the overall efficiency of the system.

Communication breakdowns contribute to an additional 9.1% of human performance failures, particularly in instances where outputs are not effectively conveyed to the pilot in a manner that aligns with standard cockpit communication protocols. Effective human-AI interaction requires that the IA not only generate optimal route suggestions but also communicate them in a way that pilots can easily process and integrate into their workflow. In cases where AI outputs are ambiguous or contradict previous system recommendations, pilots may become confused about how to prioritize the new information. This issue is further compounded when multiple actors, such as ATC and AOCC, are involved in the decision-making loop, requiring pilots to make route recommendations consistent with external flight clearance updates and operational constraints.

UC2- Liability Analysis

When examining liability risks in the context of UC2, the previously established assumptions remain applicable. In this scenario as well, the AI system introduces an additional and distinct layer to the pilot's decision-making process, which, in case of error or failure, may potentially result in legal implications not only for the individual operator, but also for the deploying organisation and the system.

Due to the level of maturity of the solution, product liability primarily falls on AI developers and technology providers, who are responsible for ensuring that the AI system functions reliably, provides accurate hazard assessments, and incorporates adequate transparency mechanisms. If the IA provides a misleading route recommendation that leads to an avoidable weather-related incident, liability may be assigned to the system provider if it is demonstrated that the error arose due, *inter alia*, to flawed algorithmic design, insufficient testing, or inadequate risk calibration. Under European Union regulations, including, in addition to the PLD, the AI Act and EASA guidance, AI developers must implement robust testing protocols, continuous validation of decision-making accuracy, and clear communication of AI system

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

limitations to mitigate legal exposure. Compliance with this requirement may be facilitated through the application of consensus-based industry standards such as EUROCAE ED-324.

Organisational liability applies to deployers (e.g., aircraft operators), who are responsible for the implementation, operational use, training, and monitoring of AI-based decision-support systems. If a pilot is inadequately trained in using the IA and subsequently s/he fails to override a hazardous route recommendation, liability could shift to the airline if it is determined that insufficient training contributed to the misjudgment. Aircraft operators must also ensure that their operational protocols clearly delineate when and how AI recommendations should be validated and executed, overridden, or deferred.

Pilot liability remains a critical issue, as flight crews retain ultimate decision-making authority and are expected to cross-check recommendations against real-time environmental conditions, ATC clearances, and operational constraints. If a pilot blindly follows an erroneous AI recommendation without performing independent verification, they may be held responsible for negligence, failing to exercise due diligence in risk assessment. This places a burden on pilots to actively engage with AI decision-support tools, ensuring that their final decisions are informed by both automated insights and human expertise.

UC2 - Mitigations

Rather than being confined to individual KPAs, the proposed mitigations reflect systemic challenges identified across safety, human performance, and security, ensuring a holistic response aligned with AI regulatory requirements.

For developers and producers, the emphasis is first placed on the integrity and diversity of training, validation, and testing data. To ensure that the IA can generalise effectively across a wide spectrum of operational scenarios, it must be trained on datasets encompassing diverse weather conditions and real-time aviation constraints. Data pipelines must be secured against corruption and degradation, both accidental and malicious, through long-term data management strategies and robust cybersecurity protections. Specific attention must be paid to the risk of data poisoning and adversarial examples, which can compromise the IA's ability to detect hazards or correctly rank routing alternatives.

Accuracy, robustness, and lifecycle cybersecurity must be built into the AI model architecture from the earliest design stages. This includes documenting all

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

performance metrics, implementing system redundancies, and equipping the IA with mechanisms for detecting and correcting internal inconsistencies. Transparency is equally critical. Developers must define clear information protocols that allow deployers and pilots to interpret how input data is processed and how routing options are generated. System capabilities and limitations should be explicitly stated, with real-time monitoring features embedded in the HMI to flag unexpected deviations in system behaviour.

A second, equally important layer concerns explainability. Developers must ensure that the IA's reasoning process is intelligible, not only in testing environments but in real-world, high-stakes conditions. This requires interface designs that present prioritised outputs in ways that pilots can quickly interpret, even under cognitive stress. The inclusion of shared interpretation tools, manual verification procedures for route alternatives, and user feedback loops will be key to improving system usability, preventing automation bias, and reinforcing pilot agency.

From the deployer perspective, the focus shifts to operational integration, infrastructure readiness, and personnel training. Data quality management systems must be implemented in line with developer guidance, encompassing not only the accuracy of inputs, but also the hardware, storage, and maintenance infrastructure that supports IA operation. This includes deploying protective measures against data corruption and ensuring periodic system audits and updates to preserve functional resilience.

Training emerges as a critical mitigation layer. Pilots must receive dedicated instruction not only in how to operate the system but in how to interpret its recommendations and verify its outputs. This includes scenario-based simulations, preferably evidence-based according to ICAO Doc 9995, replicating adverse weather and time-pressured re-routing situations, with a strong emphasis on collaborative decision-making, interface fluency, and emergency response protocols. The training must also address the behavioural risks of overreliance and under-reliance, equipping operators to maintain active oversight and critical judgment when interacting with the system.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

UC3 - Digital Assistant for UAM Coordination (DUC)

The Digital Assistant for UAM Coordination (DUC) is designed to enhance real-time decision-making and airspace management in Urban Air Mobility (UAM) operations. It functions as an AI-driven support system for UAM Coordinators, facilitating the prioritisation of emergency traffic, real-time airspace adjustments, and dynamic hazard mitigation. The validation of UC3 is based on two operational scenarios, each illustrating distinct challenges related to AI-assisted UAS Traffic Management (UTM) in urban environments.

The first scenario focuses on an in-flight medical emergency, where a passenger onboard an unmanned vertical take-off and landing (VTOL) capable aircraft (i.e. type #2 certified operations) becomes unresponsive, requiring immediate rerouting to the nearest hospital vertiport. The DUC's role is to prioritize the emergency flight, ensuring airspace clearance, trajectory optimization, and coordination with ground medical responders. The system must rapidly assess conflicting flight paths, dynamically update airspace restrictions, and provide the UAM Coordinator with precise real-time recommendations to facilitate a safe and efficient diversion. The primary risks in this scenario revolve around route optimization failures, discrepancies in airspace clearance protocols, and remote pilot hesitancy in executing automated recommendations.

The second scenario involves an urban fire near a vertiport, necessitating a partial U-space closure to prevent unmanned aircraft from entering a hazardous airspace region. The DUC must analyse real-time environmental data, determine the extent of the closure, and guide UAM Coordinators in rerouting impacted flights while minimizing traffic disruptions. The challenge in this scenario lies in the reliability of geofencing recommendations, the timeliness of airspace restriction enforcement, and potential misalignments between AI-driven hazard assessments and human operator decision-making.

UC3- Safety Assessment

A structured safety-critical event analysis of UC3 (see Report UC3, Annex A) identified three primary risk categories: Lack of Preparedness, Unpredictable/Inconsistent AI Decisions, and Overreliance/Under-reliance on AI Recommendations. These systemic risks affect decision-making reliability, operational efficiency, and emergency response effectiveness.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Lack of Preparedness was the most significant risk factor in the vertiport fire scenario, accounting for 52% of safety failures, and contributed to 30% of failures in the medical emergency scenario. This risk primarily stems from insufficient personnel training, inadequate IA validation mechanisms, and gaps in communication protocols between IA and human decision-makers. In the medical emergency scenario, delays in prioritizing the VTOL emergency diversion could be due to uncertainty in interpreting route modifications. The lack of structured emergency response procedures for IA-assisted air traffic prioritization resulted in delays in medical transport approvals, increasing the risk of prolonged response times and potential negative health outcomes for the passenger. In the vertiport fire scenario, the absence of real-time synchronization between AI-driven hazard assessments and operator decision-making led to the delayed implementation of airspace geofencing. The system could fail to immediately classify the evolving fire threat, causing delays in aircraft rerouting and increased exposure of UAM traffic to hazardous conditions.

Unpredictable AI behaviour accounted for 70% of failures in the medical emergency scenario and 45% in the vertiport fire scenario, highlighting inconsistencies in system-generated recommendations and misalignment with human operator expectations. In the medical emergency scenario, the DUC's route prioritization model could display inconsistencies in ranking alternative landing sites, with suboptimal vertiports being selected over closer, more suitable locations. Additionally, DUC system may occasionally fail to account for fluctuating airspace congestion, generating rerouting recommendations that conflict with pre-existing flight paths and required manual corrections by the UAM Coordinator. In the vertiport fire scenario, the DUC's hazard detection algorithms could miscalculate fire progression patterns, leading to premature or delayed airspace restriction decisions. The IA also may experience difficulty in distinguishing between temporary obstructions (e.g., smoke drift) and persistent hazards (e.g., fire proximity to landing zones), resulting in inconsistent airspace closure recommendations.

Overreliance on AI was a less frequent but still critical risk, accounting for 3% of failures in the vertiport fire scenario and a smaller but notable presence in the medical emergency scenario. This risk may be manifested in two distinct ways. In some cases, UAM Coordinators can show excessive confidence in accepting the DUC's recommendations, failing to cross-check system outputs before making operational decisions. In other cases, operators may disregard recommendations due to previous instances of unreliable or misleading system outputs, leading to delayed or inefficient responses. The latter issue underscores the need for outputs to be transparent,

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

interpretable, and properly contextualized to maintain operator trust and encourage balanced human-AI collaboration.

UC3- HP Assessment

The HP analysis of UC3 identifies four key performance degradation areas that directly impact the UAM Coordinator's ability to make timely and accurate airspace management decisions. The most critical HP challenges include trust calibration, interface usability issues, shared situational awareness discrepancies, and communication inefficiencies. These factors influence operator workload, decision reliability, and the effectiveness of human-AI coordination in emergency scenarios.

Trust calibration emerges as the most significant HP challenge, particularly in the medical emergency scenario. The reluctance of UAM Coordinators to immediately accept rerouting recommendations results in delays in emergency flight prioritization, potentially leading to negative health outcomes for passengers requiring urgent medical attention. The root cause of this trust deficit is the AI's previous inconsistency in hazard detection and rerouting logic, where some flights were diverted unnecessarily, and others were incorrectly flagged as non-urgent. This repeated exposure to incorrect system outputs causes operators to second-guess DUC decisions, increasing their reliance on manual verification, which significantly slows down decision-making during time-sensitive scenarios. The trust imbalance leads to a paradoxical situation where insights are either disregarded due to past failures or followed without verification due to operator fatigue, both of which introduce avoidable operational risks.

Interface usability issues significantly hamper the UAM Coordinator's ability to process recommendations, particularly in the vertiport fire scenario where multiple emergency alerts are generated simultaneously. The DUC may fail to properly prioritize flight-critical warnings, leading to a scenario where operators must manually filter alerts based on their perception of urgency rather than relying on an automated triage system. This inefficiency increases cognitive workload, forcing operators to dedicate additional time to processing raw IA outputs instead of executing critical decision-making tasks. When airspace congestion is high, the interface design shortcomings force the UAM Coordinator to alternate between multiple data feeds, which contributes to delays in geofencing hazardous areas and increases the risk of unauthorized aircraft entering restricted airspace before restrictions are fully applied.

Shared situational awareness discrepancies between the DUC and the human operator further exacerbate operational inefficiencies. The AI system continuously updates its

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

hazard assessment models based on real-time data inputs, but operators often lack full visibility into how and why AI-driven risk assessments change over time. This problem is particularly present in the vertiport fire scenario, where DUC-generated airspace updates may not always align with the UAM Coordinator's understanding of fire progression patterns. The absence of real-time synchronization between human decision-making processes and risk assessments could lead to delayed implementation of no-fly zones, exposing UAM aircraft to dynamic hazards that could have been mitigated with more precise coordination between human and AI agents.

Communication inefficiencies between the DUC and human operators introduce additional barriers to effective decision-making. The IA does not always clearly indicate its level of confidence in its hazard classification and rerouting recommendations, leading to uncertainty regarding the reliability of its suggestions. The IA may fail to properly escalate critical rerouting requests, requiring manual intervention to confirm the severity of an emergency before flight prioritization is granted. The lack of clear explanation mechanisms for AI decisions forces operators to spend additional time validating system outputs, reducing overall response efficiency in high-tempo scenarios.

UC3 - Mitigations

The most critical areas of improvement involve refining AI decision-making consistency, improving interface usability and implementing structured operator training programmes.

To enhance AI reliability and decision consistency, system developers must improve the training, validation, and testing datasets used for hazard detection and flight prioritization models. AI models should incorporate machine-learning techniques capable of self-correcting past decision inconsistencies, ensuring that future rerouting recommendations remain predictable, explainable, and aligned with human decision-making expectations. In addition, real-time calibration mechanisms should be implemented, allowing the AI system to continuously assess its own decision accuracy and adjust confidence levels based on historical performance trends. This will enable the DUC to provide human operators with an accuracy rating for each recommendation, allowing for more informed decision validation and execution.

To address interface usability deficiencies, the DUC's alert management system must be redesigned to improve the prioritization and categorization of alerts. Critical flight risks should be automatically ranked based on severity, probability, and required

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

response time, ensuring that UAM Coordinators do not need to manually filter AI outputs in real-time during high-stress scenarios. Furthermore, visual representations of geofencing decisions should be improved, with a more intuitive interface that allows operators to clearly distinguish between temporary flight restrictions, dynamic hazard zones, and long-term airspace closures. These enhancements will reduce cognitive workload and allow for faster decision execution in emergency airspace management.

To mitigate trust calibration issues, structured training programmes should be introduced to help operators develop a more refined understanding of AI-assisted decision-making. These training modules should focus on scenario-based learning, exposing UAM Coordinators to simulated emergency events where they must interpret and validate rerouting decisions under time constraints. The goal is to reinforce best practices for assessing recommendations, ensuring that operators neither blindly trust nor automatically reject system outputs without sufficient verification. Additionally, training should incorporate adaptive learning techniques, where operators receive feedback on their decision validation strategies, helping them develop a more nuanced understanding of when outputs should be followed, cross-checked, or overridden based on operational context.

To address shared situational awareness discrepancies, risk assessments should be continuously synchronized with human operator workflows, ensuring that UAM Coordinators receive real-time updates on system rationale and the underlying data sources that influence decision adjustments. This can be achieved through interactive explainability features, where operators can query the AI system for additional context behind a hazard classification or rerouting recommendation. By providing operators with a clearer understanding of risk assessments, response efficiency can be improved, and decision misalignment can be minimized.

To resolve communication inefficiencies, the DUC's alert escalation protocol must be designed to ensure that critical flight interventions are properly prioritized and flagged. In the medical emergency scenario, this means ensuring that high-priority rerouting recommendations are automatically escalated for immediate action, reducing unnecessary delays in flight diversion approvals. In the vertiport fire scenario, this means establishing a standardized confidence rating system for hazard classifications, ensuring that UAM Coordinators can assess the reliability of geofencing decisions before enforcing airspace closures.

In general, to promote a clear accountability regime, regulatory bodies must define structured liability frameworks that outline the responsibility distribution between AI developers, operational stakeholders, and human operators. AI system providers

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

should be held accountable for ensuring the accuracy, transparency, and explainability of system-generated recommendations, while UTM service providers and VTOL operators should be responsible for personnel training, procedural standardization, and compliance with safety regulations. UAM Coordinators must retain final authority in emergency decision-making, but their accountability should be contingent on their ability to interpret and act upon recommendations within an established regulatory framework.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

UC4 - Intelligent Sequence Assistant (ISA)

The Intelligent Sequence Assistant (ISA) is designed to support Air Traffic Control Officers (ATCOs) by optimising the sequencing of arriving and departing aircraft at single-runway airports. The system aims to reduce ATCO workload, enhance situational awareness, and improve operational efficiency in air traffic management. ISA integrates machine learning and deep learning algorithms to process real-time air traffic data, generating automated sequence recommendations that assist controllers in making more efficient runway utilization decisions.

ISA functions as a decision-support system rather than an autonomous controller, meaning that ATCOs retain full operational authority over air traffic sequencing. The system continuously updates its recommendations based on evolving flight conditions and provides explanations for each sequence adjustment, ensuring that ATCOs remain informed about why specific aircraft are prioritized or repositioned. Despite its intended benefits, the introduction of ISA into high-stakes air traffic control environments presents significant safety, human performance, and liability challenges, which are assessed through a detailed evaluation of safety-critical scenarios.

UC4 - Safety Assessment

The safety analysis of UC4 (see Report UC4, Annex A) reveals critical vulnerabilities in AI-assisted sequencing decision-making, system reliability, and human-AI coordination in tower and remote tower operations. Three primary categories of safety risks have been identified: Lack of Preparedness (16%), Unpredictable/Inconsistent AI Decisions (87%), and Overreliance on AI (33%). It is important to note that some tasks may provide evidence of different safety risks simultaneously, so the percentages are not cumulative.. These risks contribute to suboptimal aircraft sequencing, increased ATCO workload, and potential runway incursions, particularly during high-traffic or complex operational conditions.

Lack of preparedness presents operational risks associated with inadequate ATCO training, failure to anticipate ISA system dependencies, and the absence of structured procedures for engaging and disengaging ISA in different air traffic conditions. ATCOs who are not sufficiently trained in the nuances of sequencing logic may misinterpret ISA-generated recommendations, resulting in incorrect clearance issuance, suboptimal spacing between aircraft, and increased aircraft taxi delays. The most concerning aspect of the lack of preparedness is the tendency of ATCOs to disable the ISA system entirely during high-traffic periods, reverting to fully manual sequencing

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

without a structured transition plan. This abrupt disengagement increases cognitive workload, disrupts sequencing continuity, and forces controllers to make high-stakes decisions without the support of AI-driven optimizations.

Unpredictable and inconsistent AI behaviour is the dominant safety risk, representing 87% of identified failures, and introduces significant hazards related to system reliability, unexpected sequencing updates, and non-transparent decision logic. ISA is expected to function as a predictive tool that optimizes runway usage and aircraft spacing, but repeated cases of incorrect prioritization, delayed updates, and system reactivation failures create substantial operational hazards. The most severe safety failures occur when ISA-generated sequences do not reflect real-time traffic conditions, leading to aircraft receiving incorrect sequencing positions or ATCOs being forced to intervene and manually re-sequence aircraft under time pressure. In the worst-case scenario, an incorrect sequence update that is not caught by the ATCO could result in a loss of separation event, missed approach conflicts, or excessive go-arounds, increasing fuel consumption and disrupting airport operations.

Failures related to ISA reactivation after system disengagement pose a major risk to air traffic control continuity. ISA may fail to correctly update aircraft sequences upon being re-engaged, generating outdated sequencing lists that do not reflect changes in aircraft positioning or new arrivals in the airspace. This type of failure forces ATCOs to manually cross-check each sequencing recommendation against real-time traffic surveillance data, delaying clearance issuance and increasing the likelihood of sequencing errors. Additionally, in cases where ISA unexpectedly overrides an ATCO's manually adjusted sequence, the controller may issue instructions based on their own expectations, unaware that ISA has recalculated a different sequencing order, resulting in potential aircraft misalignment and inefficiencies in runway occupancy.

Overreliance on AI contributes to 33% of safety failures, highlighting the risk of automation complacency in tower operations. ATCOs who may habitually accept ISA-generated sequencing updates without verifying them against real-time flight conditions expose air traffic operations to significant safety risks, including sequencing errors that could lead to aircraft clearance conflicts or taxiway congestion. One of the most concerning trends in overreliance cases is ATCOs failing to override ISA-generated outputs even when discrepancies are observed, due to increased cognitive dependency on the system's logic and a reluctance to challenge automated recommendations. The risk of blind acceptance of outputs is particularly high when ISA provides priority sequencing for certain aircraft types without an immediately

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

transparent justification, leading to unexpected operational disruptions and inefficient use of available runway capacity.

UC4- HP Assessment

The HP analysis of UC4 identifies four primary cognitive and operational challenges affecting ATCOs' ability to effectively engage with ISA-generated sequencing recommendations. In particular, out of the total 6 identified HP issues, loss of trust accounts for 5 instances, comprising approximately 50% of the total. Training-related concerns are identified in 3 instances, representing 30%. Communication issues are noted in 1 instance, constituting approximately 10%. Shared situational awareness issues are identified in 1 instance, comprising about 10%.

Trust calibration emerges as the most severe HP issue relating to inconsistent confidence in ISA-generated outputs due to prior exposure to erroneous sequencing recommendations. When controllers encounter unexpected ISA behaviour, such as incorrect prioritization of departing flights or failure to update sequencing based on real-time traffic flow, they may develop a conditioned response to manually verify or override updates, slowing down operational throughput. This issue is exacerbated by ISA's failure to provide real-time explanatory reasoning for sequencing changes, leaving controllers uncertain about the validity of automated updates and hesitant to act upon them without additional verification steps.

Interface usability presents significant obstacles to ISA adoption, particularly in situations where the system generates multiple sequence modifications in rapid succession without clearly differentiating priority-level updates from minor sequencing adjustments. ATCOs must manually sort through overlapping sequencing suggestions, leading to delays in clearance issuance and increased cognitive workload. The lack of distinct visual hierarchy in ISA's HMI layout could force controllers to spend additional time verifying sequencing outputs, reducing the system's intended efficiency benefits.

Shared situational awareness discrepancies arise when ISA modifies sequencing orders without adequately communicating the reasoning behind each update, causing ATCOs to operate with an incomplete understanding of current air traffic prioritization logic. This misalignment increases the risk of sequencing misinterpretation, as controllers may proceed with clearance issuance based on their mental model rather than the output, creating conflicts in aircraft movement coordination.

Training deficiencies are directly linked to operational errors, as many ATCOs are not adequately trained on ISA disengagement and re-engagement protocols, leading to scenarios where controllers disable the system without understanding the procedural

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

risks of reverting to manual sequencing in high-traffic environments. The absence of structured AI training modules that emphasize when and how to override ISA-generated sequencing recommendations leads to inefficient human-AI collaboration, increasing both workload and sequencing errors.

UC4- Liability Analysis

Liability issues in UC4 affect three primary entities: technology developers (as AI providers), air navigation service providers (ANSPs, as AI deployers), and individual ATCOs. Each group carries distinct responsibilities regarding the reliability, deployment, and human oversight of the ISA system.

Firstly, AI providers may face product liability risks due to unexpected system behaviours related to technical robustness and operational transparency issues. If the AI system generates erroneous sequencing recommendations that lead to incidents related to runway congestion, safety breaches or airspace mismanagement, for example, and these are attributable to system malfunction or failure, providers could be held accountable for algorithmic flaws and insufficient validation procedures. Likewise, liability may arise if the human-machine interface design or the information provided for proper system use fails to offer adequate support for effective human oversight. In such cases, the provider could be held responsible for failing to ensure that air traffic controllers (ATCOs) can understand the logic behind sequencing updates and intervene when necessary.

ANSPs assume organizational liability for ensuring that ATCOs are properly trained, that ISA deployment is accompanied by structured operational protocols, and that ATCOs are prepared and well-trained to override or validate sequencing recommendations when necessary. If an ANSP fails to provide adequate training and ISA integration protocols, liability may shift to the organization for neglecting to ensure operational safety standards.

ATCOs retain professional liability for their decisions in air traffic management. If a controller blindly follows a sequencing recommendation that leads to a runway incursion or miscoordination between arriving and departing aircraft, they may be held accountable for failing to exercise due diligence in validating system outputs.

UC4 - Mitigations

The mitigation strategy developed for UC4 reflects a comprehensive and system-wide approach. These measures are rooted in the cross-cutting findings of the safety,

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

human performance, and liability analyses and are tailored to the current level of system maturity.

For developers and producers, the first layer of mitigation focuses on the quality of training, validation, and testing data. To ensure that the IA performs reliably in a wide range of air traffic control scenarios, developers must integrate diverse and representative traffic data into the AI model. In parallel, redundant safety validation procedures must be applied during both development and deployment to capture potential errors in trajectory calculations and conflict resolution under degraded conditions. AI models must also be designed with long-term robustness and cybersecurity in mind. This includes transparent documentation of model architectures and performance metrics, continuous error tracking mechanisms, and safeguards against malicious attacks such as data poisoning or adversarial inputs, which could compromise the accuracy of AI-supported decisions.

To enhance system reliability and mitigate operational risks, ISA's decision algorithms should be validated and tested to reduce inconsistencies in sequencing recommendations. Outputs could include confidence ratings, allowing ATCOs to quickly assess the reliability of each sequencing update and prioritise their validation efforts accordingly. Moreover, ISA's HMI must be designed to visually highlight critical sequencing modifications, ensuring that priority updates are immediately distinguishable from minor adjustments. Improving the interface's real-time update display logic will reduce cognitive overload and enhance ATCO efficiency in managing AI-assisted sequencing operations.

Data management strategies must further include mechanisms to prevent degradation over time, addressing both data integrity and system retraining needs. Real-time dashboards, alerts, and performance monitoring should be incorporated into the system to detect abnormal behaviour or output deviations, ensuring that both developers and end-users can identify anomalies before they lead to operational disruptions.

Crucially, explainability is treated as a core functional requirement, not a secondary design goal. Developers are responsible for integrating explainability into both the model's internal logic and its external interface design. HMIs must be engineered to enable ATCOs to correctly interpret AI outputs and detect automation bias or unexpected behaviour. This includes intuitive visualisations, prioritised alerts, and feedback mechanisms to capture user decisions and reasons for disregarding AI recommendations. The design of these interfaces must support not only real-time

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

oversight but also retrospective review, facilitating operator learning and system improvement.

From the deployer's perspective, mitigation begins with ensuring the operational data pipeline's integrity. This involves adherence to the developer's specifications regarding data collection, labelling, aggregation, and retention, along with the provision of sufficient computational and maintenance resources. Cybersecurity defences must be embedded in all data-handling operations, including mechanisms to detect and mitigate external threats to data confidentiality and system functionality.

To support long-term system performance, deployers must also establish data and quality management systems capable of maintaining accuracy and operational resilience over time. This includes simulation-based testing, scheduled updates, and scenario-based validation to ensure the IA remains responsive to real-world complexity and changes in the air traffic environment. Periodic assessments of algorithm behaviour under stress, unexpected traffic patterns, and degraded system conditions are essential for reinforcing trust and functional continuity.

Finally, training for ATCOs and other system users plays a pivotal role in closing the gap between AI system capabilities and operational use. Training programmes must be specialised and scenario-driven, exposing controllers to the range of AI-supported outputs and decision pathways. These must also include scenario-based modules that simulate sequencing conditions, ensuring that controllers develop strong competencies in interpreting ISA-generated recommendations and detecting potential system anomalies. The training should encompass manual override techniques, AI validation strategies, and structured ISA engagement protocols to minimise sequencing disruptions caused by system failures or manual disengagement decisions. Operators must be able to confidently interpret the IA's outputs, integrate them into collaborative decision-making environments, and override or challenge recommendations when required. A clear understanding of the AI's limitations, its intended operational role, and the expected human-in-the-loop procedures is essential to prevent overreliance and ensure high-quality decision-making in time-constrained environments.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

UC5 -Airport Safety Watch (ASW)

The Airport Safety Watch (ASW) is designed to support predictive decision-making for airport safety teams by analysing historical aviation data. The system aims to enhance incident reporting, improve operational risk identification, and provide actionable safety insights to mitigate emerging hazards.

ASW is structured around two primary operational scenarios. The first scenario, Data-Driven Insights, focuses on analysing historical safety records to provide medium-term risk reduction strategies. The second scenario, Proactive Risk Prediction, enables ASW to forecast daily operational risks and suggest real-time mitigation strategies.

Currently, the development focus is on Scenario 1, with London Luton Airport (LLA) serving as the primary test environment. The main users include airport safety staff responsible for daily safety operations, while secondary users include Safety Stack stakeholders such as airlines, NATS (National Air Traffic Services), and Ground Handling Service Providers.

The system is intended to function as a continuous monitoring tool, currently updating on a weekly basis but with plans for real-time updates in the future. ASW processes multiple data sources, including weather conditions, aircraft movement logs, and human performance reports, providing a comprehensive risk assessment framework. Safety personnel interact with ASW via a desktop interface, while mobile alerts are available for operational staff. ASW functions as a decision-support tool rather than an autonomous decision-maker, ensuring that human operators retain responsibility for interpreting and validating safety insights.

UC5 - Safety and Security Assessment

Unlike the previous Use Cases, the assessment of UC5 (ASW) adopts a different methodological approach due to the nature of the system and its underlying processes. In this case, the analysis did not follow the KPAs structure applied elsewhere, as the categories used in the prior assessments did not fully capture the specific characteristics and risks associated with ASW.

Instead, the joint safety and security evaluation was structured around the system's data processing workflow, which represents the core operational logic of the AI system. Starting from this workflow, the assessment team applied a HAZOP-inspired approach to systematically identify potential hazards and security vulnerabilities

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

across each data processing phase. This method allowed for an integrated examination of both safety and security issues, which emerged jointly during expert discussions and scenario-based reviews.

The assessment is therefore articulated into four critical steps: *Data Collection, Data Ingestion/Pre-processing, Data Computation, and Data Visualization/Interpretation.* These represent the key phases where AI-driven safety analysis may introduce operational and systemic risks. The following sections provide a detailed account of the main safety and security concerns identified at each stage of the ASW process.

For further methodological details and supporting examples, please refer to the UC5 Report listed in Annex A.

Step 1: Data Collection

The first step in ASW's safety analysis process involves collecting and structuring safety incident reports. Multiple challenges arise at this stage, including missing information, inconsistent reporting formats, and delayed data entry.

One of the most significant risks at this stage is insufficient evidence or missing contributory factors in incident reports, leading to underrepresented hazards in AI risk models. If critical incident details are omitted from reports, ASW may fail to recognize key risk patterns, reducing the effectiveness of predictive safety alerts.

Another challenge involves fragmented reporting of contributory factors, where similar incidents are logged differently depending on the safety team member reporting the event. This inconsistency leads to variability in insights, causing discrepancies in how risks are assessed and prioritized

The failure to capture relevant safety violations presents another major concern, as many minor operational hazards go unreported. If past safety records do not fully represent all operational risks, ASW's training data remains incomplete, leading to reduced predictive accuracy.

Finally, delayed incident data entry further compromises risk prediction. If incidents are not logged promptly, the AI system may fail to incorporate recent safety concerns into its risk assessments, increasing the likelihood of outdated insight

Step 2: Data Ingestion and Pre-processing

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Once safety reports are collected, ASW ingests and processes them to prepare for AI-driven analysis. This phase introduces risks related to excessive data ingestion, security vulnerabilities, and compromised data quality.

A major safety risk occurs when ASW ingests excessive amounts of data, generating too many low-priority alerts. This contributes to information overload and decision fatigue for airport safety personnel, reducing trust in ISA outcomes.

Security vulnerabilities arise when confidential safety data leaks during the ingestion process, increasing the risk of unauthorized access to sensitive airport operations data. Similarly, data poisoning—where malicious actors manipulate AI training data—could lead to corrupted safety insights and misleading risk assessments

Dropped data during ingestion presents another hazard, as missing records can skew trends, leading to inaccurate risk prioritization. If ASW fails to ingest and retain complete datasets, its predictions may be based on partial or incorrect information, limiting its reliability.

Step 3: Data Computation

The AI-driven computation phase applies machine learning models to extract risk insights. However, this phase introduces risks related to incorrect feature extraction, flawed data transformation, and incomplete risk assessment metrics.

One of the most critical computation failures occurs when ASW extracts incorrect features from safety reports, leading to misclassification of incident severity and incorrect mitigation recommendations. If key risk factors are omitted or misinterpreted, ASW may generate misleading safety insights, compromising airport risk management strategies.

Logical programming errors can further distort ASW's ability to assess risks accurately, causing misalignment between recommendations and actual safety needs. If the system miscalculates the probability of an incident recurrence or underestimates risk severity, safety personnel may deprioritize critical hazards that require immediate attention

Step 4: Data Visualization and Interpretation

The final stage of ASW's risk assessment process involves presenting information and recommendations to safety personnel. Usability challenges in the system's interface directly impact how operators interpret AI recommendations.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

One of the most common issues is misleading insights from excessive data warnings, leading to an overload of alerts that obscure genuinely critical safety risks. This phenomenon, known as the "Christmas Tree Effect", results in decision fatigue, delayed response times, and reduced confidence in AI-driven safety alerts.

If ASW generates delayed or incomplete safety alerts, operators may fail to act on emerging risks in a timely manner, increasing the likelihood of preventable incidents. Additionally, biased insights stemming from incomplete data representation can cause AI recommendations to favour certain risk factors over others, leading to an imbalanced allocation of safety resources.

Liability Assessment in AI-Driven Aviation Safety (UC5)

This section examines the allocation of liability within the operational context of an AI-powered Aviation Safety Warning (ASW) system (UC5), in light of the results obtained by the SHS analysis and the peculiar use of HAZOP in this context.

The integration of ASW introduces specific considerations, yet it largely operates within established aviation liability frameworks. As an AI deployer, the airport managing organization (e.g., London Luton Airport - LLA) bears primary legal responsibility for airport safety and security. Accordingly, it is mandated to ensure regulatory compliance, data integrity, and the integration of ASW insights within its established Safety Management System (SMS). This organizational accountability aligns with the principle that the entity introducing a product or service is responsible for its safe and compliant operation.

The liability analysis, leveraging the operative risks identified by the HAZOP methodology—which in this UC was applied distinctively to the ASW development pipeline—specifically focused on mitigating potential product liability risks *by design*.

As a premise, according to the new EU PLD regime of 2024, the ASW could be classified as a "product". Consequently, its developers and providers may incur legal responsibility for defects, inaccuracies, or failures. Liability assessment for ASW identified the following potential legal risks:

- **Design Defects:** Arise if ASW's algorithms inadequately incorporate critical risk factors or misinterpret safety trends, leading to erroneous recommendations. This could establish developer liability for harm resulting from flawed AI risk assessment models.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

- **Manufacturing Defects:** Pertain to software malfunctions, data processing errors, or cybersecurity vulnerabilities that compromise ASW's operational integrity. The provider may be liable for damages caused by such system failures.
- **Warning Defects:** Occur if ASW's recommendations lack transparency or fail to provide adequate user instructions, potentially leading to overreliance or misinterpretation. Developers could be liable for insufficient warnings regarding system limitations.

These risks were identified according to the new broad definition of product 'defectiveness' (Dir. (EU) 2024/2853) and considered the new requirements established by the new EU AI regulations for aviation (Reg. (EU) 2024/1689 and EASA guidance on AI). A proactive adherence to relevant legal frameworks in light of the risks detected by the SHS-L approach can contribute to a more robust development of the solution, mitigating by design the potential legal risks. In this regard, to address product liability exposure, recommendations emphasized robust documentation of design choices for transparency in both development and deployment. Additionally, implementing robust and resilient mechanisms for error detection was deemed crucial.

Building on this background, complementary considerations were made about the potential issues in terms of organisational liability and professional liability for the deploying organisations and human operators.

As anticipated, airport operators, including their safety management teams, are legally responsible for ASW's proper implementation, oversight, and operational integration. Negligence in safety management, such as inadequate personnel training on ASW correct usage, deployment without sufficient validation, or mismanagement of AI-based recommendations, could result in liability claims if AI-driven misjudgements contribute to operational hazards or aviation incidents. Implementing the ASW, airports shall thus ensure the system complements, rather than replaces, human decision-making; undergoes continuous evaluation to prevent reliance on inaccurate or outdated models; and is integrated with clear organizational guidelines defining its role in safety planning.

Accordingly, individual airport safety personnel acting as end-users should always monitor, validate and critically implement ASW recommendations. While potentially professional negligence claims could be conceivable if their use or misuse of ASW directly leads to preventable incidents, in principle their liability exposure remains

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

limited. This is because ASW functions as an assistive tool within a comprehensive safety framework, reducing the likelihood of direct legal repercussions unless gross negligence or wilful misconduct is demonstrated.

UC5 - Mitigations

To ensure the safe, effective, and compliant deployment of the Airport Safety Watch (ASW) system, the HAIKU framework outlines a comprehensive set of mitigation strategies structured across several interdependent domains: data quality and reporting culture, IA validation and monitoring, data ingestion and cybersecurity, human oversight, and training.

At the foundation of mitigation is the improvement of incident data quality and the cultivation of a robust, transparent reporting culture. Accurate, structured, and complete safety data is essential for generating meaningful insights. Developers are encouraged to enhance incident reporting templates by incorporating detailed contextual elements, especially concerning human factors and the actions of flight crews. Equally important is the promotion of consequence-free reporting practices among operational staff. By clearly demonstrating how ASW-derived insights contribute to safety improvements, the system fosters trust and incentivises comprehensive, timely data submission — addressing concerns about potential reputational or procedural consequences.

A second pillar of the mitigation strategy concerns the validation and continuous monitoring of ASW performance. Establishing baseline comparisons — for instance, through before-and-after assessments of incident rates following ASW-guided interventions — enables stakeholders to quantify the impact of AI-supported recommendations. In parallel, post-deployment monitoring ensures the timely detection of performance drift, feature degradation, or misclassification trends. To maintain operator trust and system integrity, real-time confidence scoring mechanisms should be integrated to help users evaluate the reliability of system outputs.

A third critical area involves the refinement of data ingestion processes and the establishment of clear data relevance criteria. Filtering protocols must be optimised to reduce the ingestion of redundant, low-value, or noisy data, which may otherwise lead to alert fatigue, false positives, or cybersecurity vulnerabilities. Developers should clearly define which data types, severity levels, and operational contexts are pertinent, thereby improving consistency, system responsiveness, and risk classification accuracy. These measures should be coupled with advanced encryption standards

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

and multi-layered authentication protocols to safeguard sensitive safety information from unauthorised access, data poisoning, or system manipulation.

Cybersecurity is recognised as a central enabler of overall system reliability. End-to-end protection strategies are needed to counter threats such as data corruption, denial-of-service attacks, or internal misuse. These include secured communication channels, audit trails for data handling, anomaly detection mechanisms, and fail-safe recovery procedures — all of which help preserve operational continuity and protect stakeholder trust.

Human oversight remains essential to avoid automation complacency and ensure that ASW functions as a decision-support tool rather than a decision-maker. Operators must retain the ability to verify, challenge, and override AI-generated recommendations. To this end, structured training and guidance materials are fundamental. These include user manuals, contextual on-screen prompts, interface-specific tutorials, and digital walkthroughs that clarify complex functionalities — such as the Zoom-In module — and prevent cognitive overload or bias reinforcement.

Finally, stakeholder empowerment is supported through progressive, scenario-based training programmes. These are designed to enhance AI literacy, foster interpretative skills, and equip airport safety personnel with the competence to translate system outputs into operationally relevant actions. Training modules cover both real-time and post-event decision-making contexts, enabling users to engage critically with predictive visualisations, prioritisation logic, and forecasting outputs. Together, these measures ensure that ASW implementation is underpinned by both technological robustness and human adaptability, maximising its role as a transformative enabler of aviation safety.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

UC6 - COVAID

COVAID is designed to enhance passenger navigation and health safety within airport environments by minimizing exposure to high-risk areas and optimizing movement patterns in real time. The system leverages a combination of IoT sensors, artificial intelligence, and real-time data analytics to provide dynamic route recommendations, ensuring passengers can navigate the airport efficiently while avoiding areas of high congestion and potential health hazards.

The system integrates multiple data sources, including real-time occupancy sensors, LIDAR-based crowd monitoring, air quality measurements, and passenger movement analytics, creating a comprehensive risk assessment model. These inputs allow COVAID to evaluate environmental conditions and passenger flow trends and generate personalized routing recommendations that prioritize safety, efficiency, and accessibility.

Passengers interact with COVAID through a dedicated mobile application, where they receive optimized route recommendations, real-time safety alerts, and congestion updates. The system dynamically adapts to changing environmental conditions, ensuring that passengers can modify their travel paths to avoid high-risk zones. Airport operators access COVAID through an analytical dashboard that provides an overview of passenger distribution, congestion hotspots, and emerging risk factors, enabling them to make informed decisions on resource deployment, crowd management, and safety protocols.

UC6 - Safety Assessment

The safety assessment of COVAID focuses on systemic vulnerabilities in passenger navigation and health risk mitigation within high-density airport environments. The identified risks are classified into three primary categories: lack of preparedness, unpredictable or inconsistent AI decisions, and overreliance on recommendations. These risks have direct consequences for passenger well-being, disease transmission prevention, and the overall effectiveness of the system

Lack of preparedness represents the most significant safety risk associated with COVAID, accounting for 66.7% of critical safety incidents. The primary concern arises when passengers fail to engage with the system effectively, either due to unawareness of its functionality, refusal to use it, or lack of confidence in navigation recommendations. This creates a gap between the system's intended impact and its

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

actual effectiveness, as passengers who do not utilize the app are left vulnerable to congestion and exposure to high-risk areas.

Passengers neglecting to use the app may either not download it at all or disregard warnings, leading to avoidable exposure to crowded or high-risk zones. Some passengers may deliberately opt out of COVAID's recommendations, preferring traditional navigation methods such as following standard airport signage or relying on personal observation. This behaviour diminishes the effectiveness of the system's risk mitigation strategies and creates inconsistencies in crowd control efforts.

Unpredictable or inconsistent AI decisions account for 22.2% of the identified safety risks. These failures stem from errors in routing recommendations, where passengers may be guided towards unintended locations due to incorrect occupancy calculations, data processing errors, or system lag in updating congestion status. This creates scenarios where passengers unknowingly walk into areas of increased health risk, despite having followed suggestions.

There is a critical risk of AI-driven miscalculations, particularly when real-time passenger flow data does not accurately reflect dynamic changes in movement patterns. If sensor-based congestion estimates are delayed, or if environmental variables such as ventilation and air quality are not factored into risk assessments, the system may underestimate exposure risks and provide misleading recommendations.

Overreliance on AI represents 11.1% of the safety risks associated with COVAID. Passengers who develop excessive trust in routing suggestions may follow them without critically assessing environmental conditions, leading to potential conflicts between AI recommendations and real-world observations. In cases where the AI fails to update congestion risks in real time, passengers who assume COVAID is always accurate may unknowingly enter hazardous areas despite visible warning signs or crowd build-up. Passengers relying exclusively on directions without independently verifying environmental conditions present an additional concern. If the system provides an outdated routing recommendation that conflicts with an ongoing airport emergency or a sudden surge in passenger flow, the absence of human judgment may lead to congestion issues or unexpected safety hazards

UC6 - Security Assessment

The security assessment of COVAID was conducted using SecRAM 2.0, a security risk assessment methodology derived from SESAR, which provides a structured approach

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

to identifying threats, vulnerabilities, and necessary security controls. Due to limitations in the current TRL of COVAID, the assessment was restricted to asset identification and impact analysis, with detailed vulnerability mapping and threat likelihood estimation remaining incomplete.

The primary assets identified in the security assessment include users' mobile phones, which serve as the main interface for interacting with COVAID, and the server hosting the IA system, responsible for processing passenger data and generating routing recommendations. Supporting assets include the data collected by the system, which consists of user preferences, movement patterns, and personal location information. While this data is crucial for the system's functionality, it also introduces privacy and data security risks if improperly managed or exposed.

The primary cybersecurity threats identified in the assessment include Distributed Denial of Service (DDoS) attacks, which could render COVAID inaccessible to passengers, and unauthorized access to users' mobile devices, which could lead to data breaches or manipulation of routing recommendations. Malicious actors could exploit security gaps in the system to redirect passengers into unsafe areas or compromise user privacy by collecting location data without consent.

One of the most concerning risks involves manipulation of routing recommendations, where attackers could intentionally interfere with COVAID's algorithms to redirect passenger flows. This type of attack could increase congestion in sensitive airport zones, disrupt crowd control efforts, and compromise overall passenger safety. Ensuring data integrity and system robustness against adversarial threats remains a critical security challenge.

COVAID also presents potential legal and ethical challenges related to privacy, physical integrity, and data protection. Since the system collects and processes personal location data to generate customized recommendations, compliance with GDPR and aviation cybersecurity regulations is essential. Unauthorized tracking or unintended exposure of personal data could result in regulatory non-compliance and legal consequences for airport operators and system developers

Due to the high variability in passenger engagement levels and the dependence on multiple external data sources, the system's overall exposure to cyber threats remains difficult to quantify, necessitating ongoing security monitoring and iterative refinements to strengthen defensive mechanisms against potential attacks.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

UC6 - Mitigations

The mitigation strategies for UC6 focus on addressing identified risks in safety, security, and passenger engagement, ensuring COVAID operates reliably while maintaining regulatory compliance and user trust.

Passenger engagement must be enhanced through awareness campaigns that promote system adoption and educate users on its benefits. The system should incorporate mechanisms to explain routing recommendations in clear, interpretable terms, reducing scepticism and increasing compliance with navigation paths. Integrating feedback loops will allow passengers to report inconsistencies, improving AI decision-making over time.

AI reliability enhancements should prioritize algorithm refinement, real-time data validation, and redundant verification layers. Improving the accuracy of occupancy estimates and risk calculations will reduce false positives and incorrect routing recommendations. AI transparency features should be implemented, ensuring that passengers understand why specific routing decisions are made and allowing them to override system-generated suggestions when necessary.

Security protections must focus on data encryption, access control, and cybersecurity monitoring. End-to-end encryption of passenger movement data will prevent unauthorized interception, while multi-factor authentication mechanisms will restrict system access to authorized users. Regular penetration testing and anomaly detection systems should be deployed to identify and mitigate cybersecurity threats before they impact system performance.

Legal and regulatory compliance requires strict adherence to GDPR and aviation cybersecurity directives, ensuring passenger data is collected, processed, and stored in compliance with privacy laws. Regulatory oversight mechanisms should be established, defining clear accountability structures for routing decisions to protect users' rights and ensure transparency in system governance.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

4. Comparison of the UCs

A first point of comparison concerns the completeness of the KPA coverage. UC1 was comprehensively assessed across all four KPAs. In contrast, other UCs were evaluated only on specific KPAs relevant to their system nature and TRL (Fig 3).

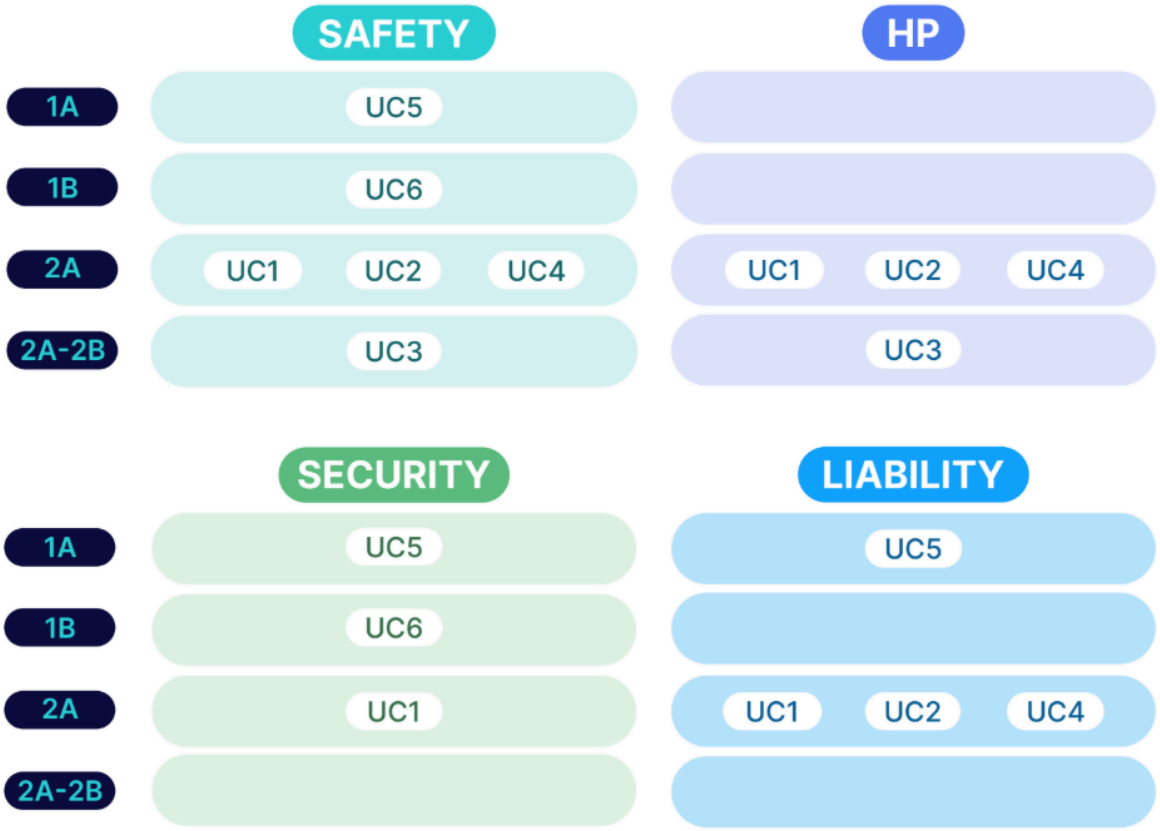

Figure 3. Overview of KPAs assessed across UCs, mapped against the corresponding EASA AI Levels.

The analysis demonstrates a clear gradient in the immediacy and severity of safety consequences associated with AI support. In UC1, where the assistant interacts directly with a pilot experiencing startle or surprise, safety risks are defined by the system's ability to detect and respond to cognitive degradation in real-time. Any malfunction, misclassification, or delay in support could immediately jeopardise flight safety, making system accuracy and response speed critical. Conversely, UC2 OLIVIA shifts the focus to decision support in route optimization, where AI errors — such as

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

mis-prioritising hazards, proposing unsafe reroutes, or failing to account for meteorological threats — can degrade safety over time rather than instantaneously. Here, the risks are less acute but can accumulate, especially in complex weather or traffic conditions. UC3 presents a hybrid context where the assistant directly influences UAM flight paths in urban airspace, including emergency rerouting. The primary safety risks are not rooted in direct physiological impairment, as in UC1, but in procedural clarity and real-time coordination. Errors in hazard detection (e.g., misjudging fire proximity) or route prioritisation can create conflicts between emergency flights and routine traffic. In UC4, where the assistant supports ATCOs during critical operational states, safety hinges on the system's ability to provide accurate, context-aware support without overwhelming the controller. Safety risks emerge when the AI generates incomplete or misleading recommendations under time pressure, potentially leading to loss of separation or missed conflict detection. UC5 takes a more strategic approach, focusing on the long-term improvement of airport safety management through historical data analysis. While the consequences of inaccurate AI insights are not immediately catastrophic, they can result in systemic safety degradation over time if incident trends are misinterpreted or critical risks are overlooked. UC6, in contrast, directly impacts public safety by offering routing suggestions in passenger terminal environments. The risks here are primarily indirect but significant — poor user engagement, misinterpretation of AI advice, or compromised data integrity could lead to crowd congestion in vulnerable areas, particularly in health-sensitive contexts.

The following sections will present a comparative assessment of the UC, based on individual analyses (see reports in Annex A) and comparing them across the four SHS-L KPAs.

4.1 Safety Considerations

The safety assessment across the six HAIKU Use Cases applied a structured analytical lens based on three predefined categories of risk: *(1) Unpredictable or Inconsistent AI Behaviour, (2) Lack of Preparedness, and (3) Overreliance on AI Outputs*. While each use case involved different actors, technological architectures, and operational pressures, applying this triadic framework enables a coherent comparison of how AI systems may affect safety performance in aviation. The differences observed do not stem from arbitrary variability, but rather from how AI functionality interacts with

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

human roles, automation levels, and environmental complexity. The following sections offer an integrated analysis of each risk category, highlighting similarities and divergences across use cases.

Unpredictable or Inconsistent Behaviour

The risk of inconsistent or context-insensitive AI behaviour cuts across all use cases, but its relevance varies with the system's role and its proximity to time-critical decisions. In pilot-facing applications (UC1 and UC2), unpredictability is particularly salient due to the cognitive and temporal pressures inherent in flight operations. In such contexts, even slight delays, misinterpretations of physiological or situational cues, or unanticipated logic in the assistant's recommendations may undermine the operator's situational awareness or lead to hesitation during degraded conditions. When the system fails to match the operator's mental model—whether due to poor timing, lack of transparency, or inappropriate prioritisation—the risk is amplified.

This form of unpredictability assumes different significance in UC3 and UC4, where the IAs operate within broader traffic management contexts. In UC3, the dynamic nature of UAM operations makes the assistant's ability to adapt to real-time hazard evolution and coordinate multiple stakeholders particularly safety-critical. A failure to reassess priorities quickly—such as not reclassifying landing sites or reacting to emergent congestion—may result in unsafe routing. Similarly, in UC4, the assistant's capacity to maintain operational relevance during degraded tower conditions is key. If flow management advice is delayed, insufficiently detailed, or unsynchronised with the controller's evolving strategy, the risk lies not in the AI's outright failure but in subtle misalignments that erode trust and effectiveness.

In more data-driven environments, such as UC5, unpredictability takes a latent form. Here, the assistant's value depends on its ability to infer meaningful safety signals from historical data. Systematic biases—such as overemphasising frequent but low-severity incidents—may distort operational focus, while inconsistent alert thresholds can lead to false reassurance or misdirected interventions. Thus, while not manifesting in real-time, unpredictability still poses a long-term safety risk by influencing how safety priorities are interpreted and addressed.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Lack of Preparedness

Preparedness gaps appear across use cases, affecting both system capabilities and user readiness. In UC1 and UC2, the interaction between AI systems and operators under stress or uncertainty presents a significant challenge. If the assistant is unable to anticipate the onset of cognitive degradation or tailor its support to degraded human performance, its presence may become counterproductive. At the same time, if users are not adequately trained to recognise, interpret, or override IA behaviours—especially when under pressure—then preparedness becomes a shared liability.

In air traffic control scenarios such as UC4, preparedness involves the system's ability to generate meaningful and timely contingency options under degraded conditions, and the operator's familiarity with when and how to rely on those suggestions. The challenge here lies in enabling effective human-AI coordination despite disruptions, requiring a level of procedural and cognitive readiness that goes beyond routine interaction.

UC5 introduces another layer of preparedness risk. Safety professionals may not be equipped to critically assess the AI's analytical outputs—particularly when model limitations or data skew are not readily apparent. Without structured validation mechanisms or a robust understanding of how the assistant generates its predictions, teams may rely on outputs that appear credible but lack operational accuracy. Here, preparedness is less about reaction time and more about analytical literacy and institutional integration.

Overreliance on AI Outputs

Overreliance—defined as an operator's tendency to defer excessively to AI outputs—emerges as a psychological and procedural risk across multiple domains. In cockpit scenarios (UC1 and UC2), overreliance may arise from perceived authority: pilots may increasingly expect the assistant to detect and manage complex events autonomously, reducing their own engagement or vigilance. When AI recommendations are accepted without critical evaluation—particularly under time pressure or during degraded states—the human operator becomes a passive node in the loop, and resilience is weakened.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

In UC3 and UC4, the pressure to act quickly and maintain operational continuity may similarly lead users to accept instructions without full validation, especially in cases where system transparency is limited or feedback loops are weak. In these environments, trust may be necessary for efficiency, but excessive trust introduces risk when IA assumptions do not align with real-time constraints or human expectations.

UC5 and UC6 illustrate subtler, but no less significant, manifestations of overreliance. In UC5, there is a risk that safety professionals would refer to data-driven outputs without questioning the underlying assumptions or data integrity. In UC6, the issue is compounded by the non-professional status of users: passengers are likely to interpret system guidance heuristically and act on it without reflection. Overreliance here does not stem from trust in automation per se, but from a lack of operational competence or contextual awareness. This behavioural vulnerability can become critical in situations requiring rapid adaptation, such as terminal congestion or evacuation.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

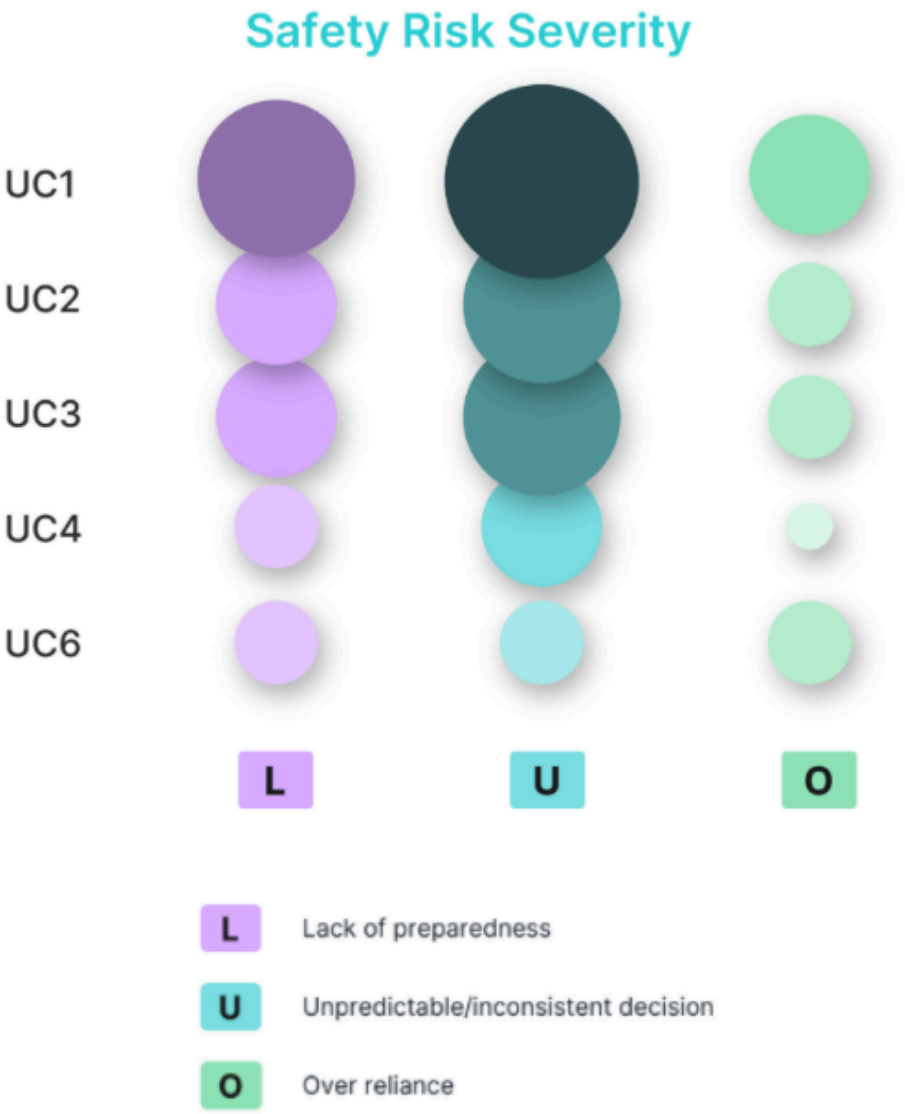

Figure 4. Overview of safety issues identified across the different UCs. The size and shading intensity of each circle correspond to the number and severity of safety issues observed: larger and darker circles indicate a higher concentration of issues.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

4.2 Human Performance Considerations

The HP assessments conducted across UC1 to UC4 reveal distinct yet interrelated patterns of how AI integration in operational aviation domains affects human cognitive performance, procedural reliability, and coordination. Using the SHS analytical categories — Interface and Interaction, Shared Situational Awareness, Communication, Trust, and Training and Operational Readiness — it is possible to extract cross-case insights and observe how specific HP risks manifest depending on the operational context and level of human-AI coupling.

Interface and Interaction issues were observed in all four use cases, but with different implications. In UC1, where the assistant is meant to support pilots recovering from startle or surprise, interface design directly affects the pilot's ability to process guidance under cognitive impairment. Misaligned timing, intrusive modalities, or poorly calibrated interaction styles risk worsening disorientation rather than aiding recovery. In UC2, where the pilot remains cognitively intact but is evaluating complex route options, interface-related risks stem from lack of clarity in how options are ranked, lack of justification for suggestions, and insufficient differentiation of hazard types, leading to interpretative delays and suboptimal decisions. In UC3, interface interaction complexity was driven by DUC generating hazard alerts and route priorities. UAM Coordinators may struggle to process recommendations when they lacked clear visual prioritization or context, particularly during emergencies. In the medical emergency scenario, operators experienced confusion when the DUC's route optimization displayed inconsistent landing site rankings or fails to highlight congestion risks, complicating decision-making. In the urban fire scenario, interface complexity may emerge when geofencing updates are not clearly distinguished between temporary obstructions (e.g., smoke) and active hazards, leading to conflicting airspace restrictions.

Shared Situational Awareness (SSA) is a critical dimension in safety-critical operations involving distributed human and AI agents. In UC1, SSA breaks down when the assistant's assessment of the pilot's mental state or task priorities diverges from the pilot's internal perception, especially when the AI intervenes too early or too late. The assistant must dynamically attune its support level to the pilot's fluctuating awareness — a capability that remains difficult to calibrate. In UC2, OLIVIA's route optimisation could be misaligned with the pilot's judgement of environmental or traffic threats, creating discrepancies in situation assessment. The IA may prioritise fuel savings over turbulence avoidance, while the pilot sees safety margins as paramount. In UC3, SSA depends on the triangulation between the different actors involved. When the assistant

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

proposes a route based on assumptions not shared by the UAM coordinator, or vice versa, divergence in expectations can lead to routing conflicts, procedural confusion, or non-compliance. UC4 highlighted a similar issue: when ATCOs receive contingency strategies that do not account for their real-time mental traffic model or sector constraints, trust in the system decreases and alignment breaks down.

Communication, both implicit and explicit, plays a foundational role in mediating human-AI collaboration. In UC1, communication challenges were centred on the assistant's ability to convey the urgency, relevance, and nature of its recommendations under high stress. If alerts or guidance are unclear or poorly timed, the pilot may be overwhelmed or fail to act. In UC2, communication issues are tied to ambiguity in hazard labelling, route justification, and system response explanations. The absence of operational explainability (OpXAI) leads to pilots second-guessing the system or bypassing its suggestions altogether. In UC3, communication breakdowns are largely procedural: the assistant's instructions sometimes clashed with the UAM coordinator's guidance. In UC4, system suggestions could arrive either too late or without enough contextual anchoring, requiring the controller to mentally translate abstract outputs into actionable steps, often during high-load periods.

Trust dynamics were a dominant concern in UC1 and UC4, and a meaningful factor in UC2 and UC3. In UC1, pilots are more likely to reject system inputs after one or more perceived failures, and more likely to over-rely in moments of cognitive overload. 70.8% of all HP issues identified in UC1 are trust-related, underscoring the volatility of reliance during degraded states. In UC2, trust is impacted by the opacity of the system's optimisation criteria: when pilots cannot understand why a route is recommended, they either dismiss it or follow it passively, both of which undermine safe decision-making. In UC3, trust varies depending on temporal reliability and consistency with instructions. Operators tend to trust the assistant when it supports their understanding of the apron environment but lose confidence when recommendations diverge from expectations. In UC4, trust issues emerge when the AI's proposed contingency measures lack traceability or fail to account for controller workload — especially in degraded modes of operation, where the stakes are high and controller confidence is already under pressure.

Training and Operational Readiness considerations are consistently identified as enabling or limiting factors in effective AI use. In UC1, training is critical to ensure that pilots understand the assistant's logic, activation thresholds, and intervention modalities. Pilots unfamiliar with the assistant's behaviour are more likely to be startled by the system itself or misuse its features. In UC2, training gaps result in

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

misinterpretation of hazard prioritization and AI route logic. Pilots often lack strategies for cross-validating the assistant's output, especially under time constraints. In UC4, controllers require training not just in how to use the assistant, but in how to interpret its outputs under degraded conditions, how to reconcile them with live sector conditions, and how to integrate AI logic into existing contingency planning workflows.

Taken together, these findings confirm that Human Performance considerations must be tailored to system role, operational context, and user expertise. UC1 highlights the need for AI systems to adapt to impaired cognitive states, providing timely, coherent support without compromising pilot agency. UC2 emphasizes the importance of aligning AI optimization logic with pilot values and constraints, requiring transparent explanations and clear communication of trade-offs. The core challenge of UC3 lies in maintaining accurate hazard awareness, route prioritization, and responsive geofencing. The DUC must continuously assess and update airspace conditions, identify and prioritize emergency flights, and guide operators in dynamically rerouting traffic away from hazards. The effectiveness of the DUC is defined by its ability to generate clear, context-aware recommendations that align with operator expectations, maintain shared situational awareness, and ensure that human operators retain the authority to validate, override, or adapt insights. UC4 stresses the importance of reducing cognitive workload and providing context-aware suggestions in degraded operational conditions, where usability and effective decision support are paramount. These diverse HP considerations highlight the necessity of tailoring AI systems to specific user expertise, cognitive load, time constraints, and human authority to ensure safe and effective human-AI collaboration.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

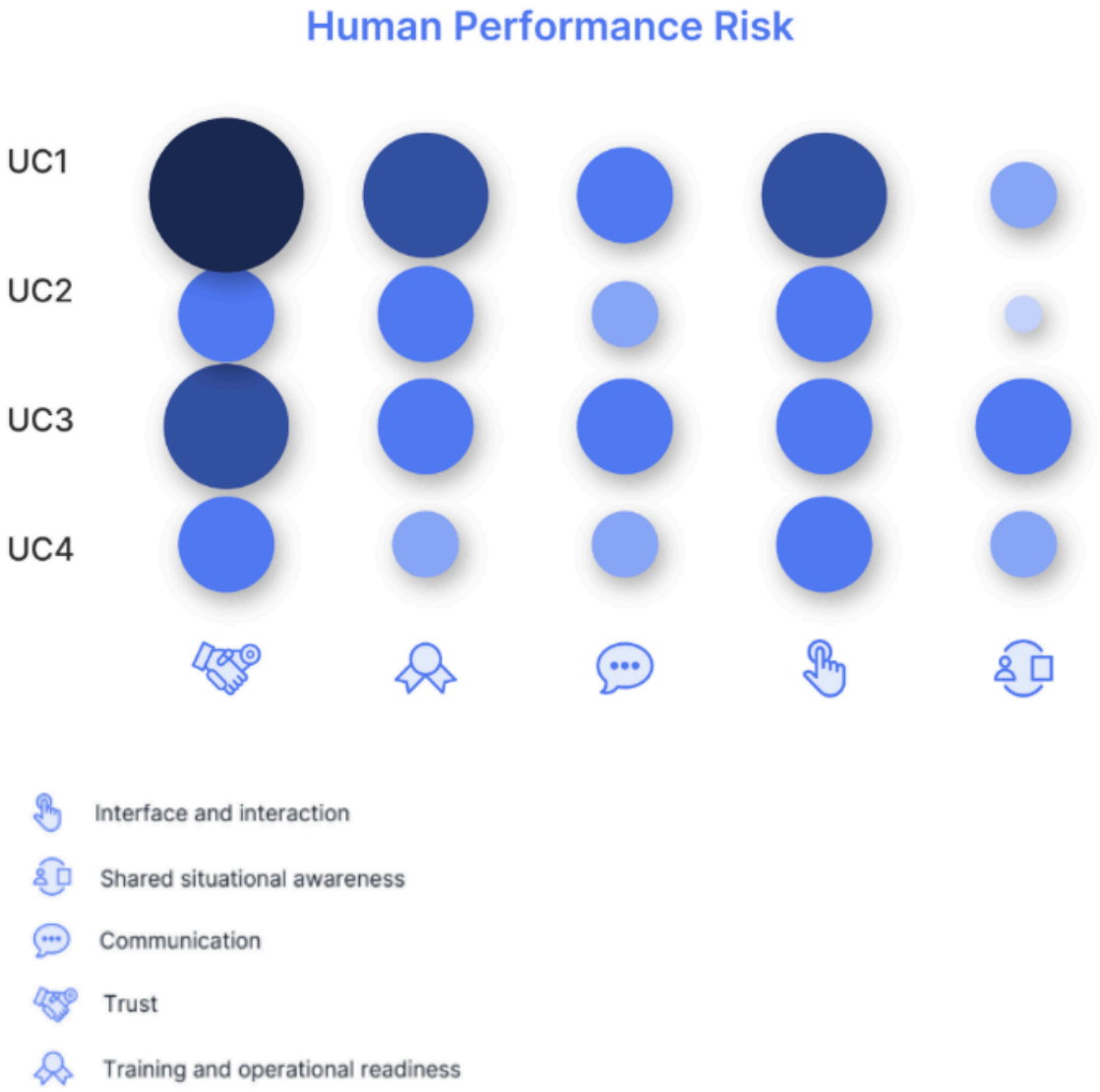

Figure 5. Overview of HP issues identified across the different UCs. The size and shading intensity of each circle correspond to the number and severity of HP issues observed: larger and darker circles indicate a higher concentration of issues.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

4.3 Security Considerations

The security assessments conducted across UC1, UC5, and UC6 collectively underscore the critical importance of embedding early, comprehensive, and context-specific cybersecurity planning into the development of AI systems, particularly in aviation and safety-critical environments. Despite the varying user profiles, system architectures, and operational scopes of these use cases, a consistent pattern emerges: while security risks are acknowledged, protective measures remain largely undeveloped or reactive. Notably, the maturity level of most systems limited the scope of security risk analysis, which in turn affected the depth and applicability of the chosen methodologies.

Given the relatively low TRLs of the evaluated systems, it was understood that a full security risk assessment could not be completed. Instead, the analysis was restricted to the identification of primary and supporting assets and a preliminary evaluation of Confidentiality, Integrity, and Availability (CIA) vulnerabilities. In most cases, the SecRAM methodology—adapted from SESAR—could not be meaningfully pursued beyond these early stages. As a result, only UC1 and UC6 underwent a structured SecRAM-based assessment. In contrast, the HAZOP methodology proved more suitable at this maturity level, yielding more detailed and actionable risk insights across the use cases.

In UC1, the assistant's security profile was shaped by its dependency on sensitive real-time physiological and behavioural data. Although operating in a secure cockpit environment, the system remains exposed to potential data manipulation or degradation, which could undermine its support function during high-stress flight phases. The SecRAM analysis identified risks related to the misinterpretation of biometric data, inappropriate system activation, and lack of authentication or anomaly detection protocols—risks that could compromise pilot situational awareness without sufficient mitigation. UC5, the AWS, presented lower immediate exposure but highlighted vulnerabilities in the integrity and confidentiality of operational safety data. The primary concern was internal misuse or ingestion of corrupted data into predictive risk models, potentially skewing safety insights. However, no formal adversarial modelling was conducted, and recommendations such as access control and penetration testing remained at the conceptual stage, with no clear evidence of implementation. UC6 posed the greatest exposure, as a mobile application processing personal location and behavioural data in a public, uncontrolled environment. The system's threat profile included DDoS attacks, unauthorised access to user devices, and manipulation of routing outputs. Despite these high-risk vectors, no advanced

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

cybersecurity controls had been implemented at the time of assessment, and key protections—including real-time validation, intrusion detection, and ethical safeguards—were still in development. The lack of formal vulnerability mapping or resilience strategies further emphasised the urgency of advancing the system's security maturity prior to deployment.

Taken together, these cases reflect the broader challenge of aligning security assessment practices with system maturity. While SecRAM provided a structured starting point, its effectiveness was limited under current TRLs (See Deliverable D7.2). The need for adversarial threat modelling, continuous monitoring, and enforceable baseline protections remains pressing. The findings reaffirm that security must be integrated from the earliest stages of AI system design—not as a reactive addition, but as a foundational layer of operational integrity, regulatory compliance, and user trust.

4.4 Liability Considerations

The liability assessment conducted across UC1, UC2, UC4, and UC5 yields uniform and consistent outcomes. At the current maturity level of the scenarios considered, the focus has primarily been on product liability risks, intending to provide recommendations that support AI providers in proactively mitigating their exposure through design-based strategies. Complementarily, the analysis also considered the organizational and professional liability risks affecting deploying organizations and human operators. These actors, indeed, will need to reconsider the current legal definitions of their roles and responsibilities when implementing and using AI-based solutions.

For the sake of clarity, the overall results of these assessments will be presented according to the different legal risks identified, drawing on the experience gained across the various UCs to outline the main lessons learned from the HAIKU project. Further details on this topic are available in D7.4 – *Recommendations for Liability by Design*, delivered in March 2025 (M31).

Starting with product liability, the recent renewal of the product liability regime (Directive (EU) 2024/2853) has introduced further uncertainty concerning the allocation of liability in operational contexts involving AI systems. In particular, the directive raises complex questions around the distribution of responsibility in systems where AI actively contributes to operational decisions. **The impact of the redistribution of authority (and associated responsibility) according to the level of**

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

AI embedded in the solutions constitutes a recurring theme that, as already observed and further discussed below, underpins and connects all the considerations explored in this analysis (references listed in Annex A).

Given that AI is increasingly deployed to replace or augment human activity—often within complex, collaborative environments—the new broad interpretation of defectiveness, grounded in general safety expectations, may expose providers to liability even in the absence of operator negligence. This is particularly relevant in teaming or advanced automation scenarios, where responsibility could shift toward the AI system itself in the event of safety failures, despite operators acting within the scope of their assigned duties.

In this regard, the initial integrated assessment of Safety, Human Factors, and Security (SHS)—including liability—conducted at an early stage of the project provided a valuable opportunity to reflect not only on operational implications, but also on broader impacts related to the characterisation of AI-enabled solutions, in line with EASA guidance on AI in aviation.

In this regard, the analysis revealed that in scenarios involving human-machine interaction classified as Level 1B: Human Cognitive Assistance (e.g., UC6), liability risks were found to be limited. At this level, AI primarily supports human decision-making, while ultimate control and accountability remain with the operator. As a result, user professional liability remains predominant, and corporate or product liability is only marginally impacted, since these systems do not introduce substantially greater risks compared to traditional non-AI tools.

In contrast, use cases featuring more advanced AI capabilities—such as Level 2A: Human-AI Cooperation (e.g., UC1, UC4)—present increased liability risks. In these configurations, indeed AI takes an active role in decision-making, potentially shifting the liability balance between human operators and the system itself. Issues of unpredictability and opacity in AI behaviour further compound product liability concerns, particularly where key operational tasks are delegated to the system.

Building on these considerations, the analysis concentrated on the most compliance-sensitive aspects of the development lifecycle, using identified critical scenarios to determine which phases—such as design, development and deployment—are most exposed to liability risk. This analysis enabled a more granular understanding of how risk is distributed across the AI development process, how it can be systematically mitigated during software and interface implementation, and how this risk information can be effectively communicated to downstream

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

stakeholders responsible for deploying the technology within their organizations. This represents a significant achievement, as it extends the scope of risk management beyond technological performance and operational usability, by also addressing the organizational and procedural challenges that may emerge over the lifecycle of the solution.

The qualitative and quantitative analysis revealed that most liability-relevant scenarios centre on technical choices, particularly those related to data preparation, software development, and interface design. However, several of the scenarios considered in the UCs also underscored the importance of informing end users about potential residual risks. When the deploying organization is involved in the system's design, such risks should be jointly assessed and addressed by both the developer (as vendor) and the purchaser. Conversely, in cases where the organization acquires a ready-made solution, the risk analysis must be handled separately, accounting for the interdependencies and responsibilities embedded within the broader system value chain.

Shifting the focus to organisational liability, this extended approach also examines how safety-related legal risks emerge throughout the operational deployment phases of AI systems, with the aim of proactively addressing potential vulnerabilities in both the implementation of new solutions and the design of related procedures.

At the organizational level, a growing set of compliance obligations is emerging—not only for AI system providers, but also for deploying entities. These obligations are generally well captured in current regulatory frameworks, which focus on ensuring accountability, traceability, and auditability[1]. However, such frameworks still fall short when it comes to clearly delineating the legal responsibilities of individual human operators. The existing definitions of roles, tasks, and competencies remain rooted in paradigms designed for conventional automation, which do not adequately reflect the evolving nature of human–AI interaction.

The implementation of AI solutions opens the door to the emergence of entirely new roles, with substantial organizational and legal implications. Early inclusion of legal and operational considerations for such novel roles proved valuable in clarifying their scope of authority and accountability. Insights from other research efforts support this approach, showing that automation often necessitates the introduction of new actors

[1] EU AI Act (Reg. (EU) 2024/1689). EASA (2023). EASA Artificial Intelligence Roadmap 2.0. May 2023 and delivered Concept Papers Issues 01 and 02.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

to manage emergent complexity and ensure system-wide safety and performance[2]. Comparisons with traditional roles such as air traffic controllers helped identify both overlapping and distinct responsibilities, offering a foundation for defining training needs, decision-making authority, and liability boundaries.

However, the operative formalization of these roles by developers and deployers organisations within existing liability and regulatory frameworks remains a challenge. A deeper understanding is needed of how new contributors affect the overall distribution of responsibility—particularly in terms of how liability is shared or reallocated across the system when errors or failures occur. By embedding legal reasoning into the design phase, the HAIKU project has demonstrated how the integration of AI into complex operational environments must be accompanied by a redefinition of human roles—both existing and emerging—to ensure clarity, fairness, and legal certainty throughout the system's lifecycle.

These potential misalignment between the operative definitions of new roles and the applicable legal regimes also reveal critical implications for the liability risk exposure of human operators. In particular, attention should be paid to potential gaps between operative and legal outlines of their role, as they are required to interact with AI-based systems in the performance of their duties, often within safety-critical contexts. These considerations highlight the need to reassess existing definitions of professional accountability in light of new forms of human-AI interaction.

The above mentioned regulatory issues create a grey area regarding operator liability, particularly in relation to the human oversight provisions established in the EU AI Act. Operators are expected to exercise discretion when using AI outputs—deciding whether to rely on, override, or suspend system actions. However, in practice, these decisions are highly context-dependent and may place undue pressure on operators, especially in safety-critical environments where AI tools are introduced specifically to enhance decision-making under uncertainty and possibly time pressure.

Generally, a potentially key concern regards the lack of preparedness in addressing inconsistent or unpredictable decisions generated by AI systems. These risks become especially pronounced in human-AI teaming scenarios, where the AI's influence on outcomes is greater than in simpler support contexts. The possibility of such failures

[2] AEON Project (H2020 - GA ID 892869) Deliverable 5.2 - Human Performance Assessment Report, August 2022. ASTAIR Project (HE - GA ID 101114684) Deliverable 5.2 - DES HE ERR ASTAIR, June 2025.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

occurring without clear warning has raised serious concerns among both implementing organizations and AI providers.

Use cases within the HAIKU project—particularly UC1, UC2, and UC4—highlighted this gap. While the AI systems considered are classified within Levels 1 and 2 of the EASA taxonomy (i.e., cognitive assistance and human–AI teaming), the level of automation introduced, especially at Level 2, necessitates a redistribution of tasks between humans and machines. This redistribution raises concerns about how responsibility will be assessed ex post in the event of failure or dispute. The problem is especially acute for actors with established legal accountability, such as pilots or air traffic controllers, who are often held responsible not only for their direct actions but also for supervising the overall operation. Under current ICAO provisions as transposed in EU common rules, such actors retain formal accountability—even when tasks are effectively delegated to AI systems functioning autonomously.

A critical challenge, therefore, lies in understanding the liability exposure of operators when they act in line with outputs or, conversely, when they override or disregard those outputs. These scenarios also raise questions about the type and adequacy of training required to prepare operators for such responsibilities under evolving legal and technical conditions.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

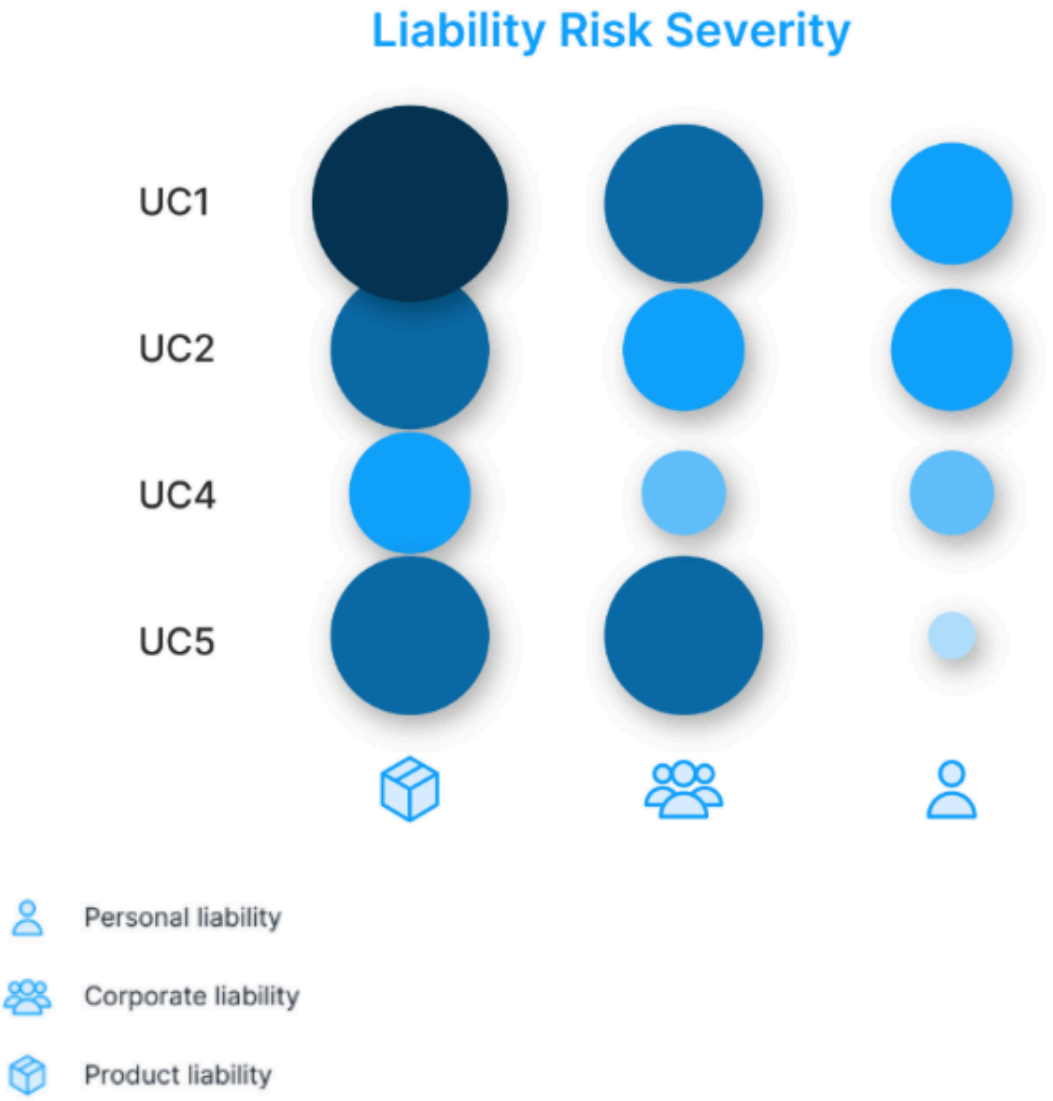

Figure 6. Overview of Liability issues identified across the different UCs. The size and shading intensity of each circle correspond to the number and severity of HP issues observed: larger and darker circles indicate a higher concentration of issues.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

5. Conclusions

The findings contained in this Deliverable D7.3 confirm that Human Performance (HP) considerations cannot be addressed through a one-size-fits-all approach, but must be carefully tailored to the role of the system, the operational context in which it is deployed, and the level of expertise of the human operator. Each use case examined illustrates that the interaction between human and IA is shaped not solely by interface design or system reliability, but by a complex matrix of cognitive demands, decision responsibilities, procedural expectations, and organisational constraints. This tailoring, however, must be based on clear, common, and legally binding rules that are agnostic to the specific process, technology, and application. The applicable rules extend beyond aviation-specific provisions to include broader acts such as GDPR, PLD, and the AI Act. Facilitating the application of these rules could involve consensus-based industry standards (e.g., EUROCAE ED-324 or ISO 27001) or Evidence-Based Training as guided by ICAO Doc 9995.

Across all four relevant use cases, it became clear that the performance of the AI system in safety-critical real time applications is inseparable from its effect on human judgement, operator confidence, and procedural control. IAs are not merely an information source or decision engine; they actively shape how operators perceive, evaluate, and act upon the operational environment. Importantly, these effects are not generic, but deeply task- and user-specific. The same system behaviour may aid performance in one domain and degrade it in another, depending on the user's role, training, and operational demands. Therefore, the different KPAs must be assessed not only for correctness of interaction but for appropriateness of support — that is, whether the AI enhances the operator's ability to perform their role, under their conditions, and within their cognitive and procedural constraints.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

References

EASA. (2023). Artificial Intelligence Roadmap 2.0 – Human-centric approach to AI in aviation (Version 2.0, March 2023). European Union Aviation Safety Agency.

EASA. (2024). EASA Concept Paper: Guidance for Level 1 & 2 machine learning applications – A deliverable of the EASA AI Roadmap (Issue 02, March 2024). European Union Aviation Safety Agency.

European Parliament and Council. (2018). Regulation (EU) 2018/1139 of 4 July 2018 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency and amending Regulations (EC) No 2111/2005, (EC) No 1008/2008, (EU) No 996/2010, (EU) No 376/2014 and Directives 2014/30/EU and 2014/53/EU, and repealing Regulations (EC) No 552/2004 and (EC) No 216/2008 and Council Regulation (EEC) No 3922/91. Official Journal of the European Union, L 212, 1–122. Consolidated version (01.12.2024): http://data.europa.eu/eli/reg/2018/1139/2024-12-01

European Parliament and Council. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Official Journal of the European Union, L 1689. ELI: http://data.europa.eu/eli/reg/2024/1689/oj

ICAO. (2006). Convention on International Civil Aviation (Doc 7300/9). International Civil Aviation Organization, Chicago, Illinois, USA.

ICAO. (2018). Annex 11 – Air Traffic Services (15th ed., July 2018). International Civil Aviation Organization.

ICAO. (2024). Annex 2 – Rules of the Air (11th ed., July 2024). International Civil Aviation Organization.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Annex A - UCs Reports

Report UC1: HAIKU D7.3 - Report UC1 .pdf
Report UC2: HAIKU D7.3 - Report UC2.pdf
Report UC3: HAIKU D7.3 - Report UC3.pdf
Report UC4: HAIKU D7.3 - Report UC4.pdf
Report UC5: HAIKU D7.3 - Report UC5.pdf
Report UC6: HAIKU D7.3 - Report UC6.pdf

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332