Haiku
Human AI teaming Knowledge and Understanding for aviation safety

Deliverable N. 7.2 - 2nd Release

Development of safety, HF and security approaches for Human Intelligent Assistance Systems

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 version of the HAIKU validation framework for Intelligent Assistants (IAs) in aviation. Developed within Task 7.2, the framework offers a structured and multidisciplinary methodology for evaluating the acceptability, trustworthiness, and compliance of AI-enabled systems operating in human-AI teaming contexts. The approach addresses four interdependent domains—Safety, Human Performance, Security, and Liability, (SHS-L)—and is grounded in existing international standards and regulatory initiatives, including the EASA AI Roadmap, the EU AI Act, and the NIST AI Risk Management Framework.

The document consolidates methods and tools developed throughout the HAIKU project to support the assessment of AI prototypes across various use cases and Technology Readiness Levels (TRLs). It includes: (i) validated procedures for safety and cybersecurity risk analysis, (ii) a novel human performance assessment framework tailored to Human-AI interaction, and (iii) a methodology to proactively address possible legal risks, considering the impact of innovation on the allocation of product, organisational, and personal liability. A dedicated section formalises the method for risk identification and prioritisation, followed by the definition of mitigation strategies aligned with stakeholder responsibilities and system maturity.

The development and iterative application of the SHS-L validation framework across multiple AI-enabled aviation use cases has led to three key findings. First, the integrated analysis across Safety, Human Performance, Security, and Liability dimensions provides substantially greater insight than siloed assessments, enabling the identification of systemic risks and interdependencies. Second, the modular and TRL-sensitive nature of the methodology ensures its scalability and relevance across a range of system maturities, from early design concepts to near-operational solutions. Third, the framework's emphasis on early inclusion of legal and regulatory considerations—such as liability attribution and explainability—supports a proactive "legal-by-design" approach aligned with emerging EU guidance.

As a result, the SHS-L framework offers a structured, adaptable, and human-centred methodology for the validation of AI systems in aviation. It is transferable to other safety-critical domains and can serve as a practical reference for developers, deployers, regulators, and certification authorities seeking to responsibly integrate AI-based assistants while preserving safety, human oversight, and accountability.

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	7.2
Deliverable Title	Development of safety, HF and security approaches for Human Intelligent Assistance Systems
Version	1.0
Status	Final
Responsible Partner	DBL
Contributors	Paola Lanzi (DBL), Nikolas Giampaolo (DBL), Elisa Spiller (DBL)
Contractual Date of Delivery	August 31st, 2025
Actual Date of Delivery	August 29th, 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	11/02/2025	Draft	Nikolas Giampaolo (DBL) Paola Lanzi (DBL) Elisa Spiller (DBL)	Outline
1.2	14/05/2025	Draft	Nikolas Giampaolo (DBL) Paola Lanzi (DBL) Elisa Spiller (DBL)	Draft
1.3	02/07/2025	Draft	Nikolas Giampaolo (DBL) Paola Lanzi (DBL) Elisa Spiller (DBL)	Draft
1.4	28/07/2025	Review	Simone Pozzi (DBL)	Review with comments
1.5	22/08/2025	Review	Ricardo Reis (EMBRT)	Review with comments
2.0	29/08/2025	Final version	Nikolas Giampaolo (DBL)	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
ANSP	Air Navigation Service Provider
ATM	Air Traffic Management
EASA	European Union Aviation Safety Agency
EU	European Union
HF	Human Factors
HAZOP	Hazard and Operability Methodology
IA	Intelligent Assistant
KPA	Key Performance Area
NIST	National Institute of Standards and Technology
OSD	Operational Sequence Diagram
SECRAM	Security Risk Assessment Methodology
SHS-L	Safety, Human Factors, Security, Liability
SOAR	State-of-the-Art Review
STPA	System-Theoretic Process Analysis
TRL	Technology Readiness Level
UC	Use Case

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 results of Task 7.2 "Acceptable Means of Compliance for Intelligent Assistants," conducted within the HAIKU project. It consolidates the development of a structured, multi-layered framework designed to support the validation of AI-enabled Intelligent Assistants (IAs) in aviation, with a specific focus on Safety, Human performance, Security, and Liability (SHS-L).

As the final and conclusive version of Deliverable 7.2, this document establishes a coherent and integrated methodology for the assessment of Human-AI teaming systems across a wide range of aviation applications and Technology Readiness Levels (TRLs). The framework incorporates the results of extensive research, iterative methodological refinement, and validation activities performed throughout the HAIKU project life cycle.

The HAIKU validation framework is anchored in the principles of human-centric design and trustworthy AI. It has been developed in alignment with leading international initiatives, such as the EU AI Act (European Commission, 2021), the EASA AI Roadmap (EASA, 2020), and the NIST AI Risk Management Framework (NIST, 2023), while responding to specific operational, regulatory, and societal challenges emerging from the application of AI in safety-critical environments.

This deliverable integrates and builds upon previous project outputs, specifically:

- D7.1 – which outlines the ethical, legal, and regulatory landscape for AI in aviation and provides a State-of-the-Art Review (SOAR) and regulatory mapping tools;
- D7.3 – which applies the methodologies defined in this document to HAIKU use cases (UCs), thereby validating the framework's effectiveness and refining its implementation;
- D7.4 – which articulates the overarching legal and regulatory conclusions, including a synthesis of liability risk assessments and compliance strategies.

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

1.2. Structure of the Document

The deliverable is organised into seven main sections and four annexes, each addressing a key component of the HAIKU validation methodology and its application:

- Section 1 – Defines the purpose, scope, and structure of the deliverable.
- Section 2 – Describes the structure and rationale of the HAIKU validation framework.
- Section 3 – Presents the methods used for scenario-based safety validation
- Section 4 – Introduces the method used to assess the severity and likelihood of the risk scenarios using a dedicated Risk Index Matrix.
- Section 5 – Details the process by which mitigation measures are formulated.
- Section 6 – Provides a summary of lessons learned and final guidance for applying the framework beyond the project.

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. Overview of the Developed Methodology

2.1 The HAIKU validation framework

The HAIKU validation framework provides a structured, multi-layered approach for the systematic assessment of AI-enabled Intelligent Assistants (IAs). It is designed to support the development, evaluation, and deployment of IAs at various Technology Readiness Levels (TRLs), ensuring that AI solutions evolve in a safe, human-centric, and compliant manner. Unlike traditional validation methods that focus primarily on technical performance or feasibility, the HAIKU validation framework takes a holistic, human-centric approach, ensuring that AI-enabled Intelligent Assistants (IAs) not only function effectively but also align with societal, ethical, and value-based considerations. By integrating different Key Performance Areas (KPAs), the framework analyses the IAs from various perspectives. HAIKU's iterative validation approach enables continuous assessment across different TRLs, supporting AI solutions to evolve safely, securely, and ensuring that IAs remain trustworthy, explainable, and human-centric by design. HAIKU validation framework is a contribution to furthering standards regarding these aspects.

The core objective of this framework is to facilitate seamless collaboration between human operators and digital assistants in safety-critical aviation environments. To achieve this, the validation framework is structured around four interdependent KPAs:

- **Safety –** Assessing the robustness of AI solutions to provide predictable, fail-safe behavior in dynamic and high-risk environments.
- **Human Performance (HP)** – Ensuring that AI systems are designed to support, rather than hinder, human operators by addressing usability, workload, cognitive alignment, explainability, and trust.
- **Security** – Evaluating cyber resilience, data protection, and system integrity to prevent adversarial attacks and unauthorized access.
- **Liability and Legal Compliance** – Clarifying legal responsibility across stakeholders, including developers, operators, and regulatory bodies, to address accountability in AI-driven decisions.

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

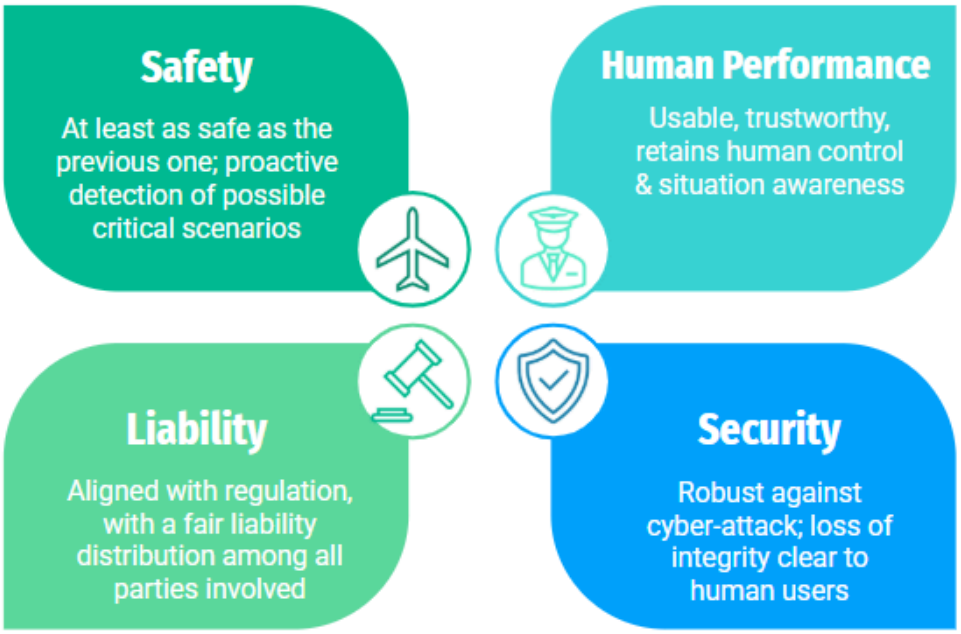


Figure 1. KPAs Dimensions

A critical strength of the HAIKU validation framework is its ability to assess AI solutions at different stages of maturity, ranging from early-stage conceptual models (TRL 2-3) to high-fidelity AI assistants (TRL 6-7). This ensures that AI-enabled Intelligent Assistants are evaluated iteratively, allowing for incremental refinements and progressive risk mitigation as they advance through development. The framework supports validation across multiple TRL stages by adapting assessment depth and complexity to the maturity of the AI system.

By enabling progressive validation, the HAIKU framework ensures that AI-enabled IAs are continuously refined based on empirical evidence, reducing unforeseen risks before full-scale deployment.

The HAIKU validation framework aligns with and builds upon existing AI governance and assurance frameworks. The framework harmonizes multiple regulatory, ethical, and operational AI principles, creating a structured approach that integrates human

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

factors, safety, security, liability, and legal compliance into AI validation. Among the key frameworks that HAIKU aligns with are:

- **The EU Framework for Trustworthy and Human-Centric AI** – This framework establishes fundamental AI governance principles aimed at ensuring that AI systems are ethical, human-centric, and aligned with societal values. It sets out key requirements such as transparency, fairness, human oversight, accountability, and safety, emphasizing the need for AI to remain explainable, and auditable across different applications.
- **The EASA AI Roadmap 2.0** – Developed by the European Union Aviation Safety Agency (EASA), this roadmap provides a sector-specific adaptation of AI governance principles tailored for aviation. It defines AI certification pathways, risk-based assurance models, and strategies for human-AI teaming, ensuring that AI systems are integrated safely into aviation operations.
- **The NIST AI Risk Management Framework (AI RMF 1.0)** – Developed by the U.S. National Institute of Standards and Technology (NIST), this framework provides a risk-based methodology for evaluating AI systems. It establishes principles covering validity, reliability, safety, security, explainability, and resilience, ensuring that AI solutions are assessed not only on their intended functionality but also on their robustness against adversarial threats, biases, and failures.

At its core, the HAIKU framework integrates AI's technical, operational, and regulatory dimensions into a single, **interrelated assessment process**, ensuring that AI solutions are evaluated not just on their functional capabilities, but also on their impact on human performance, safety, cybersecurity, and legal compliance. Unlike siloed validation methods that assess AI systems in isolation, HAIKU's multi-dimensional approach acknowledges that each Key Performance Area (KPA) is deeply interconnected, meaning that changes in one area can have cascading effects across the entire system.

2.2 Methodological Steps

The HAIKU SHS-L validation framework follows a structured, multi-step methodology designed to systematically assess AI-enabled Intelligent Assistants (IAs) in aviation, ensuring that safety, security, human performance, and liability concerns are rigorously analyzed and effectively mitigated. The methodology follows a four-stage

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

workflow (illustrated in Figure 2), ensuring a progressive refinement of risk assessment, validation, and mitigation measures. By systematically analyzing operational sequences, identifying hazards, evaluating KPAs, and implementing mitigations, the framework ensures that AI-enabled systems are operationally safe, legally compliant, and human-centered.

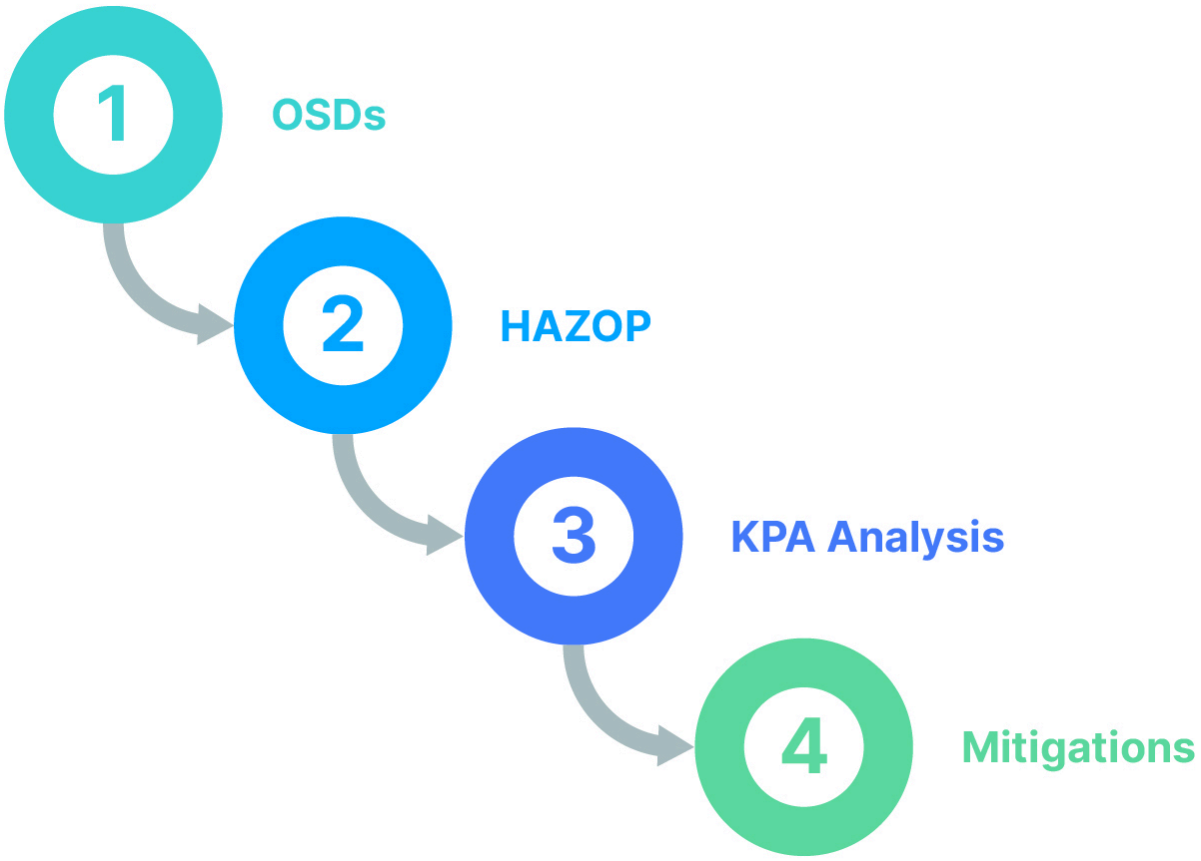


Figure 2. Methodological Steps

The first step in the process involves the use of Operational Sequence Diagrams (OSDs), an HF technique that provides a structured visualization of operational workflows, interactions, and decision points (Brooks, 1960). Within complex environments, OSDs map out how operators interact with AI systems, other crew members, and automation tools to complete critical tasks. This mapping allows analysts to identify bottlenecks, cognitive workload imbalances, and potential

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

vulnerabilities in human-AI collaboration, ensuring that workflows are optimized for efficiency, safety, and usability. By visualizing task sequences, OSDs help pinpoint areas where miscommunication, excessive cognitive demand, or system inefficiencies may arise, providing a foundation for deeper safety and performance assessments.

Building upon the insights gained from the Operational Sequence Diagrams (OSDs), the second step involves the structured analysis of potential hazards and deviations from expected operational conditions. In HAIKU, this was primarily conducted using the Hazard and Operability (HAZOP) methodology (Kletz, 1983), which proved particularly effective in identifying safety risks related to AI integration—especially where automated decision-making could lead to unintended consequences. HAZOP's structured and scenario-based format supports systematic evaluation of deviations, their potential causes and consequences, and relevant mitigation strategies under realistic operational constraints.

However, while HAZOP was selected as the primary method in this context, it is not intended to exclude the use of other safety assessment approaches. Depending on system complexity, maturity, and available engineering resources, alternative methods such as STPA (System-Theoretic Process Analysis) or other risk analysis frameworks may be equally appropriate (Sulaman et al., 2019). The essential requirement is to apply a safety assessment technique capable of producing meaningful and actionable insights that are proportionate to the system's TRL. In this regard, HAZOP is recommended as a robust option—particularly for early to mid-TRL system. The structured nature of HAZOP ensures that AI systems are evaluated under a range of operational conditions, with identified deviations assessed in terms of causes, consequences, and risk mitigation strategies.

Following the identification of potential hazards, the third step involves an in-depth analysis of critical scenarios across KPAs, covering Safety, HP, Security, and Liability, covering key risk categories (see section 2.3 and 3.2 for further details).

By integrating these multi-disciplinary analyses, HAIKU ensures that AI-enabled IAs meet the highest standards of safety, human performance, security, and legal compliance.

The final step in the validation process focuses on the development and implementation of mitigation measures, ensuring that the risks identified in the previous steps are effectively addressed. These mitigations are not limited to specific KPAs but are instead generalized, taking into account the broader systemic challenges identified across the entire analysis. By adopting a holistic, system-wide 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

mitigation strategies go beyond isolated safety or human factors improvements and instead focus on comprehensive enhancements to data management, system functionality, and human-machine interaction. This ensures that the recommendations are not only relevant to the current phase of AI development but also serve as a guide for the system's evolution, aligning future advancements with both present operational needs and anticipated regulatory requirements.

Taking into account the above, the mitigation measures are tailored to different categories of stakeholders, recognizing that AI developers, aviation regulators, and operational users each require specific interventions to effectively integrate AI into aviation systems. These measures are not static solutions but are instead designed to evolve alongside AI technologies, ensuring that as AI systems mature, their validation methodologies remain adaptive and aligned with emerging regulatory landscapes. The approach taken here emphasizes broad, system-wide improvements, ensuring that AI-enabled IAs contribute to enhanced safety, operational efficiency, and human-machine collaboration. By maintaining a continuous cycle of assessment, validation, and refinement, the HAIKU framework ensures that AI adoption in aviation is progressive, sustainable, and fundamentally aligned with the industry's safety and regulatory culture.

2.3 KPAs and Categories of Risk

The key risk categories for each KPA were defined to reflect the systemic vulnerabilities observed across the use cases. A detailed classification and description of each category is provided in Section 3.2.

From a Safety perspective, the framework focuses on three recurring risk types:

- **Lack of Preparedness,** refers to the operator's inability to effectively understand, interact with, or intervene in AI-driven processes. This is often due to insufficient training, limited exposure to AI-supported operations, or poor integration of AI into existing workflows.
- **Unpredictable or Inconsistent AI results**, which arise when the system produces false positives or negatives, misclassifies risks, or fails to respond to emerging threats. These inconsistencies may stem from biased training data, unexpected operational conditions, or model instability, among others.

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

- **Overreliance and Underreliance** on AI, reflect poor trust calibration. Overreliance may lead to automation bias and reduced human oversight, while underreliance can result in the dismissal of valid AI outputs, reducing the overall efficiency and effectiveness of the system.

The Human Performance analysis highlights factors that influence how operators perceive, interpret, and act on AI outputs. The key categories of concern include:

- **Interface Design and Interaction**, where overloaded or poorly prioritised information can impair decision-making.
- **Shared Situational Awareness,** which is compromised when human and AI systems fail to maintain a consistent understanding of the operational context.
- **Communication** limitations, particularly when AI-generated alerts are too vague, too frequent, or lack actionable detail.
- **Trust**, including mistrust, complacency, and inconsistent reliance on AI recommendations.
- **Training gaps**, reduce operator readiness and contribute to errors in both routine and high-stress scenarios.

Finally, the liability assessment ensures that legal responsibilities are clearly delineated in accordance with aviation law and the evolving regulatory landscape (e.g. the EU AI Act). The framework distinguishes between:

- **Product Liability**, focusing on design, manufacturing, and warning defects;
- **Organizational Liability**, which involves failures in oversight, training, or integration of AI into operational procedures;
- **Personal Liability**, which addresses individual accountability in cases of negligence or mismanagement of AI system interactions.

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. Methodological Foundations of the SHS-L Assessment

This section elucidates the methodology of the SHS-L framework within HAIKU, addressing two principal aspects: the adaptation and expansion of extant assessment techniques for IAs, and the identification of pivotal risk categories derived from HAIKU UCs.

3.1 OSD and HAZOP

The HAZOP methodology is a **recommended practice** within the HAIKU validation framework, offering a structured, scenario-based approach for identifying hazards and anticipating system-level vulnerabilities during the design and deployment of AI-enabled Intelligent Assistants (IAs). Originally developed within the chemical industry to detect potential deviations in complex industrial systems, HAZOP has been adapted in the context of aviation AI to address the emergent complexities of human-AI interaction, operational transparency, and shared authority in high-risk domains (Kletz, 1983). Within HAIKU, HAZOP is applied as a forward-looking mechanism to critically assess how AI systems behave under realistic operational conditions, how they interact with human actors, and where potential breakdowns in coordination, interpretation, or control might emerge.

At this stage, the methodology calls for the development of a dedicated Operations Sequence Diagram (OSD). HAIKU's approach integrates HAZOP with OSDs, a prerequisite artefact that anchors the hazard analysis in the real-time flow of operations. The OSD represents the human-AI system, capturing the sequence of actions, decisions, and interactions occurring throughout a given operational scenario. When coupled with HAZOP execution, it will support dynamic exploration of the operational activities in their context. The OSD is not just a functional blueprint; it captures each interaction step under realistic operational constraints. It integrates a detailed breakdown of environmental and contextual conditions, the state of operator situational awareness, the internal computations and decision processes of the IA, and the communication pathways between system and user. The OSD also describes how and when explainability mechanisms are activated, what form these take, and whether the rationale behind AI decisions is effectively accessible to the user. Furthermore, it

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

specifies the dynamics of authority, distinguishing between cases in which the operator retains full control or when the AI is delegated certain autonomous functions.

3.1.1. HAZOP Guidance

The quality and relevance of the HAZOP process depend substantially on the composition of the team conducting the analysis. The session must be led by an experienced HAZOP facilitator who possesses deep domain knowledge in aviation operations as well as expertise in risk and hazard analysis methodologies. A dedicated notetaker supports the process by systematically recording all identified hazards, assessed risks, and proposed mitigations. The selection of operational experts is calibrated according to the system's TRL. At lower TRLs, these may include concept developers and domain specialists with foresight expertise, while higher TRL systems require the involvement of real-world end users—such as commercial airline pilots, ATCOs, or safety engineers—who have direct experience with the IA in simulator trials or operational contexts. Human factors specialists are essential at all stages. Participation from AI developers and data scientists becomes critical in high-TRL assessments, as they are responsible for articulating system limitations, input-output logic, data quality concerns, and known model constraints. The project lead or product owner ensures traceability between hazard identification and actual system design evolution, ensuring that the process translates into tangible improvements rather than isolated observations.

The core analytical engine of the HAZOP lies in the application of Guidewords—semantic prompts such as "More Than", "Less Than", "Other Than", and "As Well As"—used to stimulate discussion around possible deviations from expected system behaviour. These guidewords are applied across each step of the interaction sequence in the OSD, generating plausible failure scenarios where the IA may deviate from expected outputs, misclassify data, or fail to interact appropriately with the user. This generative approach explores "what-if" situations that may arise under degraded conditions, edge cases, or high cognitive stress. Rather than relying solely on empirical evidence, the HAZOP process supports safety imagination, leveraging the collective experience of the assessment team to identify subtle or emergent risks that might not be apparent through conventional testing. In particular, the guideword application allows for early detection of mismatches between AI-driven logic and human decision-making, missed synchronisations between system outputs and operator intent, delayed or confusing feedback loops, and shifts in control authority that may lead to ambiguity or misjudgement in real-world operations.

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 1. HAZOP Guidewords

GUIDEWORD	MEANING
No, None	No part of the intention was achieved (omission of an action/part of a task step)
More	Too much effect resulted from the action compared to the intention
Less	Too little effect resulted from the action compared to the intention
As Well As	The intention was successfully achieved, but other effects also resulted
Part Of	Only part of the intention was achieved by the action
Reverse	The action resulted in the opposite effect to the intention
Other Than	The original intention was substituted by another action or intention
Early	The action was conducted earlier than was appropriate
Late	The action was conducted later than was appropriate

HAZOP participants are encouraged to consider potential failure scenarios as plausible risks, even if exact likelihoods cannot yet be determined.

The final stages of the HAZOP process within the HAIKU validation framework focus on documenting findings, refining system design, and increasing AI trustworthiness through iterative improvements and regulatory alignment. After systematically analyzing human-AI interactions and identifying potential deviations from expected

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

operational conditions, all insights are documented in structured tables (see D7.3). These records categorize the identified hazards and existing safeguards within the operational concept.

3.2 KPA Analysis: SHS-L

Following the OSD and HAZOP assessments, the SHS-L KPAs analysis is conducted to evaluate safety, human performance, security and liability risks in AI-enabled aviation operations. The risk categories were defined based on a detailed analysis of the individual UCs, ensuring that each assessment reflected the unique features of the specific system under study. This structured approach provided a comprehensive evaluation of the challenges inherent in Human-AI teaming and interaction, ensuring that AI integration is assessed not only from a technical perspective but also from a human-centered and operationally relevant standpoint. The selection of Safety, Human Performance, and Security risk categories in the HAIKU framework was not arbitrary but strategically aligned with the specific challenges observed across the UCs. Importantly, these risk categories identified within the framework exhibit a high degree of transversality across the various UCs. Recurring concerns—such as lack of preparedness, overreliance on AI systems, and inconsistent or unpredictable system behaviour—were observed in multiple operational settings, revealing shared vulnerabilities in human-AI interaction irrespective of the specific application context.

Recognising these cross-cutting risks highlights the added value of applying a unified SHS-L framework across diverse scenarios. It enables the identification of underlying issues that transcend individual systems while supporting context-specific interpretations based on each UC's operational and technological characteristics. These categories provide a comprehensive, comparative, and human-centered approach to evaluating IAs, ensuring that assessments are tailored while maintaining cross-UC consistency.

3.2.1 Safety Risk Categories: Addressing Human-AI Collaboration Challenges

The safety assessment within the HAIKU framework is structured around three primary risk categories: *lack of preparedness, unpredictable or inconsistent decisions, and overreliance on AI outputs.* These categories allow for a systematic comparison of safety risks across different AI applications, ensuring that safety challenges are

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

framed within a common analytical structure while being tailored to the specifics of each UC.

Lack of preparedness is a fundamental safety risk when operators are not adequately trained to understand, interact with, or intervene in AI-driven processes. The analysis of UCs demonstrated that insufficient training on AI interaction protocols, limited exposure to AI-driven decision-making scenarios, and overestimation of AI reliability led to situations where operators are unable to react effectively in high-pressure environments. Importantly, this risk also reflects a latent explainability gap, whereby operators lack the internalised understanding required to mentally reconstruct the rationale behind AI decisions during real-time operations. As outlined in the EASA AI Concept Paper (EASA, 2021), this form of explainability is not delivered explicitly by the system in the moment, but rather depends on the knowledge acquired during training, which users must retrieve and apply to interpret system behaviour. Without adequate training on system logic and customisation choices, operators may struggle to build situation awareness and appropriately validate or challenge AI outputs during critical phases of operation. This is why Lack of Preparedness most often couples with Training Deficiencies (see 3.2.2), as the absence of structured, scenario-based training prevents users from developing the cognitive strategies needed to interpret AI behaviour and maintain effective human-AI coordination.

Unpredictable and inconsistent *AI decision-making* represents another major concern identified across UCs. AI-enabled IAs may generate false positives or false negatives, leading to misclassified risks, incorrect alerts, or failure to detect emerging threats. Such inconsistencies often arise due to variations in input data, biases in machine learning models, and unforeseen operational conditions not accounted for in training datasets. Given the high-stakes nature of aviation, even minor AI miscalculations can cascade into significant safety incidents.

Overreliance and underreliance on AI systems emerged as another prevalent issue in AI-assisted operations. The assessment of UCs revealed that operators often struggled to calibrate their trust in AI recommendations, leading to two extreme behaviors. Overreliance results in complacency, where operators accept AI outputs without verification, leading to disengagement from critical decision-making processes. Conversely, underreliance occurs when operators reject AI outputs altogether, opting for manual interventions even when AI-generated recommendations are correct. Both behaviors degrade situational awareness, reduce operational

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

efficiency, and increase the risk of human error, reinforcing the need for AI trust calibration strategies that balance human oversight with system autonomy.

3.2.2 Human Performance: Evaluating Human-AI Teaming and Interaction

The HAIKU framework includes a comprehensive HP assessment. The categories selected for HP analysis are: *interface and interaction, shared situational awareness, communication, trust, and training*.

Interface and Interaction concerns arise when system interfaces present excessive, ambiguous, or poorly prioritised information, which may hinder the identification of critical alerts, distinguishing critical alerts from routine notifications, resulting in delayed responses and increased human error rates.

Shared Situational Awareness is addressed to ensure alignment between the AI system's representation of operational states and the human operator's mental model. Discrepancies in this alignment can disrupt coordinated action and impair performance in time-critical scenarios. Misalignment between AI-generated insights and human decision-making processes can lead to misinterpretations, delays, and suboptimal responses in high-stakes aviation operations.

Communication is considered a foundational element of effective Human-AI teaming. The framework recognises that both the frequency and the quality of AI-generated messages influence operational outcomes. Inadequate communication from AI systems may result in missed critical warnings or misinterpreted instructions.

Trust is treated as a dynamic and calibratable factor. Human operators must maintain an appropriate level of trust in AI outputs—neither excessive nor insufficient—to interact effectively with intelligent systems. If operators perceive AI as unreliable or unpredictable, they may disregard system outputs and rely solely on manual interventions. Conversely, automation complacency, where operators blindly follow AI-generated recommendations without verification, can lead to unmitigated errors and a decrease in human oversight capabilities.

Training deficiencies are identified as a key operational vulnerability across UCs. Inadequately prepared operators may struggle to manage AI-assisted decision-making, interpret system outputs, and utilize AI-driven tools effectively. The integration of AI into aviation workflows requires new skill sets and cognitive

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, yet without structured, scenario-based training, operators remain ill-equipped to adapt to AI-driven environments.

3.2.3 Security Assessment: The Shift from SECRAM to HAZOP

For systems at early to mid levels of maturity, the HAIKU SHS-L framework adopts a HAZOP-inspired approach as a practical alternative to full SECRAM analysis. This choice reflects the reality that, during initial development phases, many UCs lack the architectural stability, detailed data flows, and operational constraints required for a comprehensive SECRAM-based security assessment. In such contexts, HAZOP provides a structured and adaptable method for exploring vulnerabilities in AI-assisted decision-making, focusing particularly on data integrity, model robustness, and exposure to adversarial threats.

As the analysis progressed, it became clear that the maturity level of several HAIKU UCs was insufficient to support SECRAM's detailed asset modelling and risk quantification processes. AI complexity, evolving system architectures, and limited availability of operational data presented significant constraints. To address these limitations, the security assessment was reoriented using a HAZOP-based method tailored to the data processes of each UC. This adapted approach enabled a systematic examination of how AI systems ingest, classify, and act on data inputs, highlighting risks such as data poisoning, model bias, dropped records, and information leakage.

SECRAM remains a valuable and rigorous methodology for security risk assessment and is expected to become increasingly applicable as system maturity increases. Once system components are stabilised and data handling chains are well-defined, SECRAM can be deployed to support a granular assessment of critical and supporting assets, threat scenarios, and mitigation priorities. Until then, the HAIKU framework ensures that security risks are not overlooked by applying HAZOP-based analysis as an effective interim method aligned with current TRLs.

3.2.4 The liability assessment: the Legal Case to explore legal risks building on SHS

The liability assessment relies on the application of the Legal Case (Contissa et al., 2013). This methodology is designed to support the integration of automated technologies (including AI) into complex organisations, particularly in ATM. Its purpose

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

is to address liability issues arising from the interaction between humans and automated tools, ensuring that these issues are clearly identified and dealt with at the right stage in the design, development, and deployment process.

It is worth noticing that **the Legal Case is never intended to apportion liability and blame people or the organization(s). Conversely it is aimed to promote the safety culture of the organisation** making all the actors involved aware of the liability risks associated with their roles, tasks and activities and proactively identify suitable mitigations.

In fact, the method entails the 'design according to liabilities' approach. According to this, liability exposure should be considered one of the inherent properties of the socio-technical system, alongside safety, human performance, and security. As such, it should be taken into account from the earliest phases of operational concept design.

In this way, the design can be influenced by legal risks through liability considerations, while also allowing for an evaluation of their potential impact on human performance.

The Legal Case has been designed to be flexibly applied across all the phases in a system's life cycle. Depending on the technology maturity phase, the analysis will rely on different types of background information, can be used for different purposes, either proactively or retroactively..

In the proactive scenario, the liability assessment helps support and enhance the design phase of a new operational concept or system, addressing potential legal issues that may arise from future accidents or malfunctions. In the retroactive scenario, it can be applied to existing technologies to evaluate their inherent or contextual legal risks, which may evolve over time in response to changes in the surrounding environment.

The Legal Case process consists of the following four steps:

1. **Understand context and concept**. This step involves collecting and elaborating background information about the object of the study so as to understand its socio-technical and normative aspects. The information collected concerns the operational concept itself, the context of its deployment, and the legal and regulatory aspects. This step includes the identification of the AI level of the concerned system, its impact on roles, tasks and responsibilities and a set of use cases considered relevant for the following legal 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

2. **Identify liability issues**. This step involves identifying the possible liabilities related to the object of the study and determining the associated liability risks.
3. **Address the liability allocation**. This step involves analysing the acceptability of liability risks for all stakeholders, proposing possible mitigations that may improve liability allocation, and making design recommendations.
4. **Collecting findings and Systemic Analysis**. This step presents the results of the study, highlighting the liability issues associated with the object of study and the ways to deal with legal risks, as well as making further recommendations.

For the purposes of HAIKU,the Legal Case has been used in a proactive way during the design phase of a new operational concept/system. The point is to be able to address possible legal issues arising in the future from potential accidents or malfunctions. Accordingly, the liability assessment has been structured on the results produced by the OSD and the HAZOP analyses, as well as on the assessment of the other KPAs encompassed by the SHS-L framework.

The results provide a high-level overview of the potential legal risks associated with the HAIKU use cases, based on the tasks outlined in the proposed sequence diagrams. **The assessment delivers indicative outcomes focused on theoretical liability exposure, without consideration of likelihood.**

The analysis covers a broad range of scenarios, including current, new, and revised tasks. It also accounts for causal dependencies among tasks—defined as the relationship between actions and their consequences. This approach allows for a deeper examination of the potential effects of miscoordination between the involved actors and systems. The primary focus is on new or revised tasks and their associated causal dependencies. Existing tasks were considered only to the extent that they are impacted by, or connected to, the innovations introduced by the HAIKU solutions.

Based on the maturity level of the UCs, t**he main liability issues addressed included: product liability risks for AI providers; organisational liability risks for AI-deploying entities (e.g., ANSPs or air carriers); and professional liability exposure for human operators (e.g., pilots and air traffic controllers).** These liability considerations were integrated into the SHS analysis, as illustrated below.

Within HAIKU, Deep Blue has used a proprietary tool to apply the Legal Case methodology, supported by a web-based solution to ensure consistent and structured liability assessments. The tool standardizes the assessment process, and guides via legal reasoning maps in the identification of potential liability risks

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

associated with each task, summarizing relevant legal conditions and precedents. Results are presented in a risk matrix showing variations in legal risk across the concept. If the innovation introduces higher risk levels than current standards, the tool also provides recommendations to mitigate potential legal issues related to design, procedures, or operational use[1].

3.2.5 SHS-L integrated analysis

Based on the illustrated approach and methodologies, the integrated analysis of the various **KPAs** was developed as shown in the following figure. Following an initial definition and HAZOP analysis of the OSD, each identified task was mapped for potential risk sources (safety, human performance, and liability), also considering the causal dependencies between the different tasks.

[1] Further details on this topic are available in D7.4 – *Recommendations for Liability by Design*, delivered in March 2025 (M31).

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

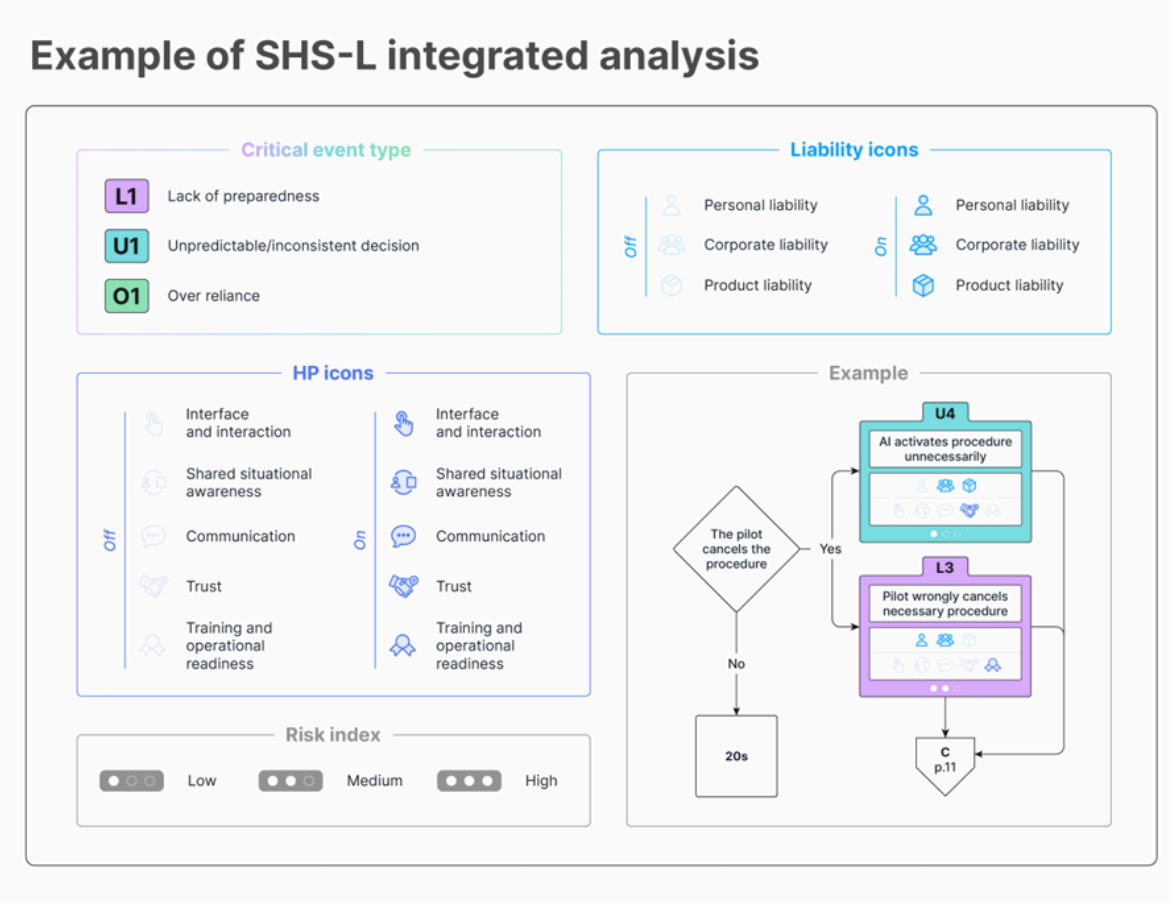


Fig 3. Example of SHS-L integrated analysis

Figure 3 offers an example of the integrated dashboard used to support the SHS-L analysis. It illustrates how the framework connects the four foundational components: KPAs, risk categories, assessment steps, and their concrete application through scenario-based evaluation. By bringing together safety, human performance, and liability dimensions within a single operational example, the figure demonstrates how safety-critical events are assessed in a holistic manner. Each event is associated with specific human performance factors, assigned a corresponding risk index, and mapped against relevant liability domains. This integrated representation enhances the traceability of risks across system components by explicitly linking each safety-critical event to associated human performance concerns, risk categories, and legal

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

responsibilities. By embedding these relationships within a unified operational scenario, the framework ensures that risk pathways can be traced from their root causes through human-system interactions to their potential safety and liability implications, but also exemplifies how cross-domain insights are operationalised to inform design improvements and targeted mitigation strategies. Ultimately, the figure encapsulates the core ambition of the HAIKU framework: delivering a comprehensive, human-centred approach to safety and accountability in the deployment of AI systems within aviation contexts.

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. Risk Identification

To ensure a structured and quantifiable approach to risk evaluation, the HAIKU framework employs a risk index matrix to assess various safety-critical scenarios. The risk index is computed by multiplying the *likelihood* and *impact* of each identified scenario.

The likelihood classification reflects the expected occurrence of each event during the operational lifespan of the AI system. Very Likely events are those anticipated to occur multiple times, with a probability greater than 1 in 100, indicating a frequent operational concern that requires significant mitigation. Likely events have a probability between 1 in 100 and 1 in 1,000, meaning they are expected to occur at least once during the system's deployment. Possible events fall within a 1 in 1,000 to 1 in 10,000 probability range, making them unlikely but still plausible. Unlikely events, classified between 1 in 10,000 and 1 in 1,000,000 probability, are rare but cannot be entirely dismissed, particularly in high-risk aviation environments . Very Unlikely events, with a probability of less than 1 in 1,000,000, are so rare that they can be considered statistically negligible.

Table 2. Likelihood Classification

Likelihood Classification	Probability Range	Description
Very Likely	> 1 in 100	Events expected to occur multiple times during the system's operational lifespan; represent frequent operational concerns requiring strong mitigation.
Likely	1 in 100 – 1 in 1,000	Events expected to occur at least once during system deployment; require systematic monitoring and mitigation strategies.

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

Possible	1 in 1,000 – 1 in 10,000	Unlikely but plausible events that may occur during extended operation; warrant preventive measures and contingency planning.
Unlikely	1 in 10,000 – 1 in 1,000,000	Rare events, but not entirely dismissible in high-risk aviation environments; should be considered in safety cases and emergency procedures.
Very Unlikely	< 1 in 1,000,000	Extremely rare events considered statistically negligible; may be monitored but not prioritised for dedicated mitigation.

In parallel, the impact classification evaluates the severity of consequences should a safety-critical event occur. Negligible impacts involve no noticeable effects on operations, such as minor delays in system synchronization that do not interfere with critical functions. Minor impacts cause slight disruptions or inefficiencies, such as brief delays in non-essential AI-driven recommendations that do not compromise overall operational safety. Moderate impacts lead to significant but manageable disruptions, such as temporary data unavailability affecting real-time decision-making, but without long-term consequences. Significant impacts introduce major operational risks, such as AI malfunctions that hinder situational awareness or response coordination, affecting the safety of human operators and system stability. Severe impacts result in catastrophic failures, including the inability to detect or respond to critical emergencies, which may lead to loss of life, severe operational breakdowns, or major damage to the aviation infrastructure.

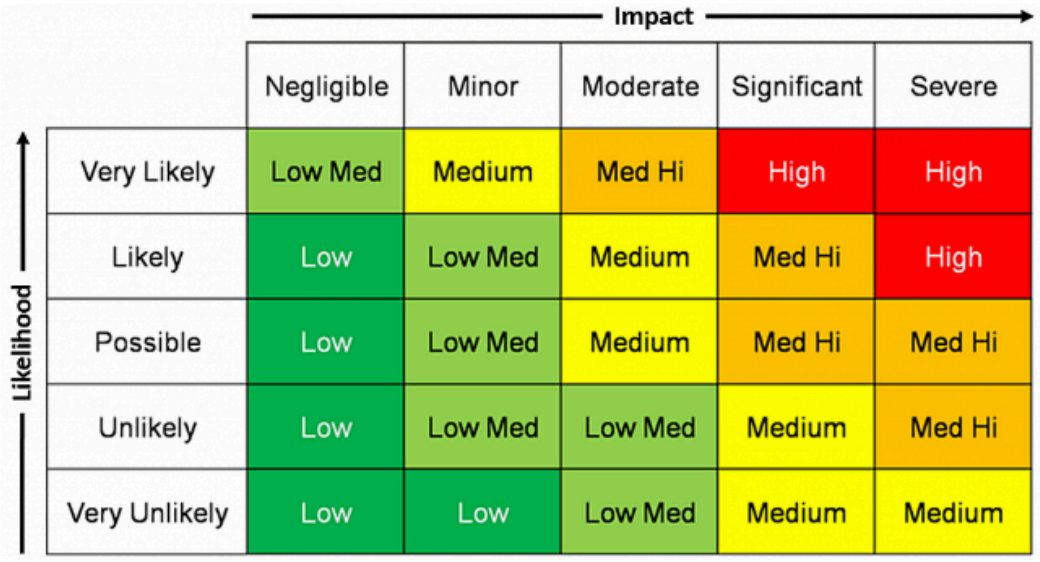
Table 3. Impact Classification

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

Impact Classification	Description
Negligible	No noticeable effects on operations; e.g. minor synchronisation delays without impact on critical functions.
Minor	Slight inefficiencies or delays in non-essential AI outputs; no compromise to overall operational safety.
Moderate	Significant but manageable disruptions, such as temporary data unavailability affecting decision-making, without lasting consequences.
Significant	Major operational risks, including AI malfunctions that impair situational awareness or coordination, with potential safety implications for operators.
Severe	Catastrophic failures such as inability to detect/respond to emergencies, possibly leading to loss of life, major infrastructure damage, or systemic breakdown.

By systematically applying this risk assessment methodology, the HAIKU framework ensures that AI-enabled aviation solutions are evaluated within a structured and measurable risk 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



Likelihood \ Impact	Negligible	Minor	Moderate	Significant	Severe
Very Likely	Low Med	Medium	Med Hi	High	High
Likely	Low	Low Med	Medium	Med Hi	High
Possible	Low	Low Med	Medium	Med Hi	Med Hi
Unlikely	Low	Low Med	Low Med	Medium	Med Hi
Very Unlikely	Low	Low	Low Med	Medium	Medium

Figure 4 - Risk Matrix

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. Mitigations

The final stage of the HAIKU framework focuses on the formulation of mitigation measures that are explicitly tailored to the roles, responsibilities, and operational contexts of the various stakeholder groups involved in AI-enabled aviation. These include system developers, regulators, operational personnel (e.g. pilots, ATCOs), and organisational safety managers. Rather than confining mitigations to specific Key Performance Areas (KPAs), the framework adopts a holistic, system-level perspective. This ensures that the proposed strategies reflect the full spectrum of insights generated through the assessment process—spanning safety, human performance, cybersecurity, and legal liability dimensions.

Importantly, the proposed mitigations are not conceived as isolated technical or procedural fixes. Rather, the HAIKU framework promotes interventions at the architectural design level, where risks can either be eliminated or mitigated in a way that delivers sustained impact across technical, operational, and organisational dimensions. This approach avoids fragile, localised solutions that may be obsolete in the face of evolving technologies, shifting operational contexts, or organisational change. Instead, mitigations are designed to address systemic vulnerabilities, ensuring resilience and long-term safety in AI-enabled aviation systems. The framework recognises that mitigating the risks associated with AI integration requires alignment with current technological capabilities and operational realities. As such, the proposed measures are calibrated to the TRL of each AI-enabled solution, ensuring their feasibility in near-term implementations while also anticipating the requirements of future regulatory developments and operational transformations.

This approach also places a strong emphasis on broad system improvements, addressing fundamental aspects such as data management, AI functionality, system reliability, and human-machine interaction.

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

6. Conclusions

The development and application of the SHS-L methodology within the HAIKU framework have provided valuable insights into both the strengths and limitations of current approaches to the validation of IAs in aviation. While the methodology has shown significant potential in supporting the development and evaluation of AI-enabled systems, it has also revealed several critical challenges that shall be addressed to enhance its maturity and applicability.

One of the most significant outcomes of this work is the **demonstration of the added value of integrated analysis compared to siloed, domain-specific assessments.** By adopting a cross-KPA perspective, the SHS-L methodology enabled the identification of systemic vulnerabilities and interdependencies that would have remained invisible, or assessed as minor, under isolated analyses. This integrative approach proved **particularly effective in highlighting areas where challenges in one domain (e.g., explainability or training in HP) were directly correlated with risks in another (e.g., liability attribution or safety oversight)**. The ability to establish these correlations allowed the project team to propose mitigation strategies that addressed multiple dimensions of risk simultaneously, resulting in more robust, systemic recommendations.

As for the challenges, the initial application of the SecRAM methodology proved problematic for lower-TRL systems, where architectural and data flows were still under definition. As a result, the methodology was adjusted, developing a tailored HAZOP approach to address AI-related data vulnerabilities and system-level security concerns. This solution demonstrates that security can indeed be integrated effectively from early stages with a flexible, context-sensitive methodology.

Another key benefit of the integrated SHS-L approach lies in its **capacity to anticipate legal and regulatory concerns early in the development process.** The traditional approach to compliance and (in case) liability typically introduces these considerations at later design stages, once system functionalities and human-machine interactions are already defined. In contrast, the HAIKU methodology introduced liability-related considerations from the outset, incorporating them into the design and evaluation of system behaviour, risk profiles, and operator roles. This proactive inclusion of legal aspects—moving beyond a purely compliance-driven perspective—has opened the possibility of embedding “legal-by-design” and “design according to liabilities”

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

approach into the development of AI systems. This is especially relevant in light of the evolving regulatory landscape (e.g., the AI Act and the EASA AI Roadmap), which increasingly demands demonstrable foresight and accountability in AI system design.

The SHS-L methodology also yielded benefits in terms of concept evolution, **supporting iterative cycles of analysis and refinement, where the assessment informed further development of the concepts**. This feedback loop contributed not only to the robustness of the methodology itself, but also to the refinement of the IAs under investigation, particularly in relation to their alignment with the EASA AI levels. For instance, in some cases, the **SHS-L analysis revealed mismatches between the proposed AI level and the expected degree of autonomy or human oversight, supporting more informed decisions about technology readiness, risk exposure, and certification pathways**.

Equally noteworthy is the **modularity of the methodology, which allowed different KPAs to be selectively applied depending on the maturity, scope, and characteristics of each UC**. This flexibility proved essential in accommodating the heterogeneous nature of the HAIKU UCs, which ranged from early-stage design concepts to near-operational systems. The ability to tailor the depth and focus of the analysis according to TRL and context ensured that the methodology remained proportionate, scalable, and operationally relevant. However, this modularity also underscores the importance of interdisciplinary collaboration. The full benefits of the methodology could only be realised when experts across domains (e.g., HF, legal, technical, operational) worked closely together throughout the process. Where this collaboration was weaker or fragmented, the integration across KPAs was less effective, and key interdependencies were more difficult to capture.

Among the specific methods applied, the combined use of OSD and the HAZOP methodology was particularly effective. The OSD provided a structured visualisation of human-AI interactions, contextualising system functions within operational workflows. HAZOP, as adapted in this project, demonstrated remarkable versatility and ease of use, even across highly diverse Use Cases. Its structured and scenario-based nature made it accessible to domain experts with limited background in Human Factors or safety analysis, thus fostering a common language for interdisciplinary discussion. This communicability greatly facilitated knowledge exchange among stakeholders, contributing to more coherent and actionable results.

Lastly, some challenges were encountered in the application of the Safety and Human Performance categories. While the selected categories proved robust and meaningful

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 most Use Cases, they were not always applicable in contexts - for instance UC5 - where the system was designed for supporting safety management activities, rather than real-time operations.

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

AI HLEG. (2020). Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment.

Brooks, F. P. (1960). Operational sequence diagrams. IRE Transactions on Human Factors in Electronics, HFE-1(1), 33–34.

Chemical Industries Association. Chemical Industry Safety & Health Council. (1977). A guide to hazard and operability studies. Chemical Industry Safety and Health Council of the Chemical Industries Association.

EASA. (2023, May). Artificial Intelligence Roadmap 2.0. Human-centric approach to AI in aviation. Cologne, Germany.

Kletz, T. A. (1983). HAZOP and HAZAN: Identifying and assessing process industry hazards. Rugby, UK: Institution of Chemical Engineers.

NIST (2023). Artificial Intelligence Risk Management Framework (AI RFM 1.0). 1-43.

SESAR Joint Undertaking. (2017). Security Risk Assessment Methodology for SESAR 2020.

Sulaman, S. M., Beer, A., Felderer, M., & Höst, M. (2019). Comparison of the FMEA and STPA safety analysis methods–a case study. Software quality journal, 27(1), 349-387.

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