Human AI teaming Knowledge and Understanding for aviation safety

Deliverable D6.3

Updated validation strategy and plan

Authors: Carl Westin (LiU)

Major contributors: Roberto Venditti (DBL), Alexandre Duchevet (ENAC), Dormoy Charles (CATIE), Anaisa Villani (EMBRT), Ricardo Reis (EMBRT), Miguel Villegas (Skyway), Barry Kirwan (ECTL), Philip Wright (ENG), Evangelos Spyrou (CERTH)

Reviewed by: Paola Lanzi (DBL), Brian Hilburn (CHPR), Vanessa Arrigoni (DBL), Simone Pozzi (DBL)

Abstract:
The HAIKU project aims to generate knowledge on intelligent assistants, and to develop AI-based intelligent assistant prototypes for six aviation-related Use Cases (UCs):

- Use Case #1 – Flight Deck Startle Response
- Use Case #2 – Flight Deck Route Planning/Replanning
- Use Case #3 – Urban Air Mobility
- Use Case #4 – Digital and Remote Tower
- Use Case #5 – Airport Safety Watch
- Use Case #6 – Airport Spreading Virus Prevention

This document provides the third deliverable in WP6 in the HAIKU project – *D6.3: Updated validation strategy and Plan*. It is considered an update of *D6.1: First validation strategy and plan*. D6.3 shall serve as a guiding document for the upcoming intelligent assistant prototype evaluations in VAL2.

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	D6.3
Deliverable Title	Updated validation strategy and plan
Version	1.0
Status	Final
Responsible Partner	LiU
Contributors	DBL, ENAC, CATIE, EMBRT, Skyway, ECTL, ENG, CERTH
Contractual Date of Delivery	August 31st, 2024
Actual Date of Delivery	October 15th, 2024
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
0.1	07-05-2024	First draft	Carl Westin (LiU)	Template
0.2	01-07-2024	First draft	UCs partners	UC contributions
0.3	02-07-2024	First draft	Carl Westin (LiU), Ricardo Reis (EMBRT)	Harmonised draft
0.4	03-10-2024	Second draft	UCs partners	UC contributions
0.5	08-10-2024	Second draft	Carl Westin (LiU), Roberto Venditti (DBL)	Harmonised draft
0.6	09-10-2024	Second draft	Paola Lanzi (DBL), Brian Hilburn (CHPR)	Review of document
0.7	10-10-2024	Final	Carl Westin	Finalisation of the document
1.0	14-10-2024	Final	Vanessa Arrigoni (DBL), Simone Pozzi (DBL)	Final quality check

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
AAM	Advanced Air Mobility
AI	Artificial Intelligence
AltMOC	Alternative Means of Compliance
AMC	Acceptable Means of Compliance
ASW	Airport Safety Watch
ATCO	Air Traffic Control Operator
ATM	Air Traffic Management
CISP	Common information service provider
CLT	Construal Level Theory
CRM	Crew Resource Management
DA	Digital Assistant
DUC	Digital Assistant for UAM Coordinator
EASA	European Union Aviation Safety Agency
FIFO	First-In-First-Out
HAIKU	Human AI Knowledge and Understanding
HAIT	Human AI Teaming
HAT	Human Autonomy Teaming
HATR	Human Autonomy Teaming Requirements
HAZOP	Hazard and Operability Study
HF	Human Factors
HITL	Human-In-The-Loop
HLR	High Level Requirements
HMI	Human Machine Interface
HPC	High-Performance Computing
IA	Intelligent Assistant
IoT	Internet of Things
IR	Implementing Rules

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

ISA	Intelligent Sequence Assistant
JRCC	Joint Rescue Coordination Center
KPA	Key Performance Area
KPI	Key Performance Index
LACC	Levels-of-autonomy-in-cognitive-control
LIFO	Last-In First-Out
LLA	London Luton Airport
LOA	Levels Of Automation
LOC-I	Loss Of Control In Flight
MbC	Management by Consent
MbE	Management by Exception
ML	Machine Learning
MOC	Means of Compliance
MoE	Measures of Effectiveness
MoP	Measures of Performance
NLP	Natural Language Processing
R&D	Research and Development
SA	Situation Awareness
SHAP	Shapley Additive Explanations
TEM	Threat and Error Management
TRL	Technology Readiness Level
UAM	Urban Air Mobility
UAS	Unmanned Aircraft System
UC	Use Case
UPRT	Upset Prevention and Recovery Training
UTM	Unmanned Aircraft System Traffic Management
VAL1	First Validation
VAL2	Second Validation
VTOL	Vertical Take-off and Landing
WP	Work Package

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

XAI	Explainable Artificial Intelligence

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 Content

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

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

Executive Summary

The HAIKU project aims to pave the way for human-centric Intelligent Assistants (IAs) in the aviation domain by developing AI enabled prototypes for six aviation-related use cases.

- Use Case #1 – Flight Deck Startle Response
- Use Case #2 – Flight Deck Route Planning/Replanning
- Use Case #3 – Urban Air Mobility
- Use Case #4 – Digital and Remote Tower
- Use Case #5 – Airport Safety Watch
- Use Case #6 – Airport Spreading Virus Prevention

WP6 has two primary objectives: to evaluate the progress of the Intelligent Assistant concepts and prototypes, and to assess the final prototypes by providing empirical evidence of their operational benefits. Deliverable D6.3, *Updated Validation Strategy and Plan*, supports these objectives by offering updated use case descriptions and detailed plans and strategies for prototype validations in the second iteration.

This report covers an update of the following activities:

- **Technology Readiness Level.** Update for all use cases, demonstrating the advancement of prototype readiness for VAL2 and the remainder of the project.
- **UC Operational concepts and descriptions.** Progress on operational requirements specification, development and validation planning specifically for the second validation (extending D4.1, HAIKU, 2024).
- **Validation framework.** Integrating all use cases into a cohesive validation strategy that transforms the D4.1 validation requirements (HAIKU, 2024), harmonised against the EASA Human-AI Teaming requirements, into specific validation objectives, activities, and metrics for the second round of validation.
- **Validation plan and strategy** outlined for each use case, with specific validation objectives, activities, and metrics based on the scope and goals of each case.

The structure of this deliverable builds and updates its predecessor D6.1: First Validation Strategy and Plan. This document specifies the Human-AI Teaming validation objectives that each use case is addressing in the second validation. A theoretical framework is provided to demonstrate how the HAIKU use cases have

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

defined Validation Objectives for the second validation in accordance with the EASA HAT framework. This framework expands upon and further refines the conceptual validation requirements outlined for each use case in D4.1 (HAIKU, 2024).

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

Introduction

This report describes the work carried out in HAIKU Task 6.6. The objective of this task is to prepare for the second validation by providing an updated version of the validation strategy and plan from Task 6.3, with IA prototypes aiming at TRL4-6. As such, the activities are similar to those of Task 6.3, with the addition of two use case-centred workshops to discuss results from the first validation (VAL1) and improvements to the validation strategy and plan and determine validation objectives that use cases will address in the second validation (VAL2).

This document presents the theoretical framework that has been used to harmonise and guide use cases' validation strategy and plans. The theoretical framework builds on the EASA Human-AI Teaming (HAT) requirements (2024) and links it to the HAIKU high-level requirements and overarching HAIKU research questions and objectives. All use cases have worked with the framework to specify their validation objectives.

This document is organised into ten sections, beginning with an executive summary in the first section, followed by an introduction in Section 2. Section 3 outlines the validation strategy and presents the Human-AI Teaming framework used to harmonise use case validation objectives, the reference EASA Validation Requirements, and an overview of the use cases validation plans and strategies. Sections 4 through 9 cover six distinct use cases. The six use cases are:

Section 4: Use Case #1 – Flight Deck Startle Response

Section 5: Use Case #2 – Flight Deck Route Planning/Replanning

Section 6: Use Case #3 – Urban Air Mobility

Section 7: Use Case #4 – Digital and Remote Tower

Section 8: Use Case #5 – Airport Safety Watch

Section 9: Use Case #6 – Airport Spreading Virus Prevention

This document provides a description of all use cases specific validation objectives, corresponding indicators and metrics and the work plan for VAL2.

All use cases follow a consistent structure with four primary subsections:

- The first subsection provides a TRL overview for each use case, highlighting how TRLs are increased in VAL2 and the remainder of 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

- The second section outlines an updated concept description, covering key R&D needs, objectives, performance targets, and high-level validation requirements.
- The third section presents the validation (VAL2) plan and strategy for each use case, with variations depending on the specific scope and objectives of each case. This includes tables outlining the validation objectives, along with the corresponding validation activities and metrics that will be investigated for each use case.
- The last section is the conclusion where some key lessons learned are reported.

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 Validation Strategy

1.1 Human-AI Teaming Validation Framework

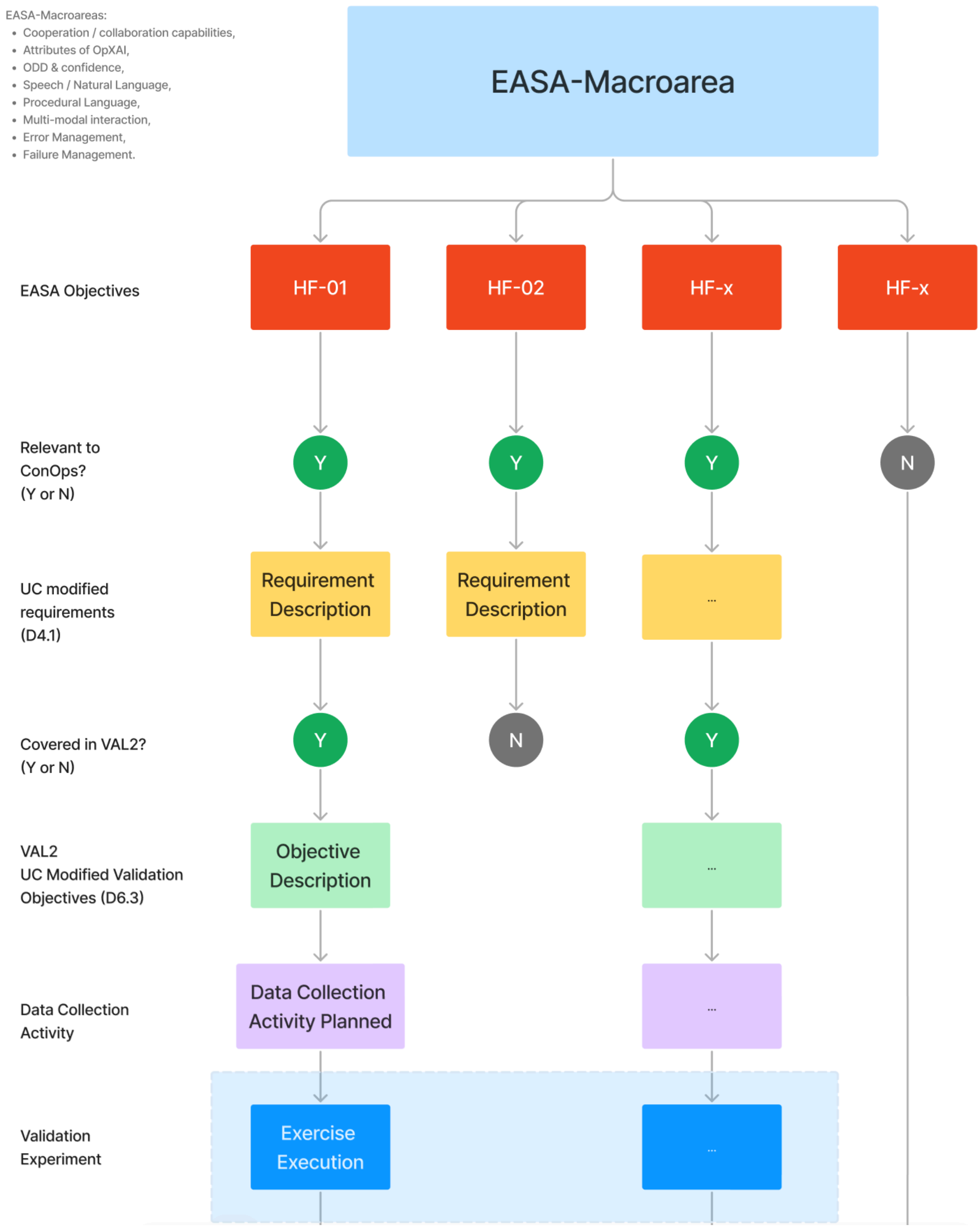
To harmonise the validation strategy across use cases, a validation framework was developed (Figure 1). Prior to this deliverable, D4.1 (HAIKU, 2024) outlined how the use cases align with the EASA HAT requirements from EASA's Roadmap for Trustworthy AI (EASA, 2020) and EASA's initial version 1 guidance for AI in aviation applications (EASA, 2023). In D6.3, use cases specify which of the EASA HAT requirements covered at an conceptual level will be explored in VAL2.

The framework shown in Figure 1 illustrates the strategy for how each use case has evolved from the validation requirements identified in D4.1 (HAIKU, 2024) to defining validation objectives, activities, and metrics for VAL2. The framework takes the shape of a flowchart, starting with the EASA HAT Requirements and further refines the conceptual validation requirements outlined for each use case in D4.1 (HAIKU, 2024). In the HAIKU project, our focus within the EASA Macroarea is on Human Factors and Explainability objectives. The next step allows use cases to answer yes/no if a requirement is addressed by the use case at a conceptual definition of the intelligent assistant (IA). A second yes/no decision gate indicates whether the validation requirement will be addressed in VAL2. If the answer is yes, the corresponding validation objectives, data collection activities, and metrics to be explored in VAL2 are specified.

For each use case, we present the results from following this process. Results are then to be consolidated at the project level for a comprehensive analysis. The results from VAL2 in answering the validation objectives, data collection activities, and metrics are then linked back to use cases' High Level Requirements (HLR) to facilitate lessons learned and conclusions. Findings from the different use cases are then used to answer the HAIKU high-level research questions and objectives.

Figure 1. *HAIKU validation 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



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

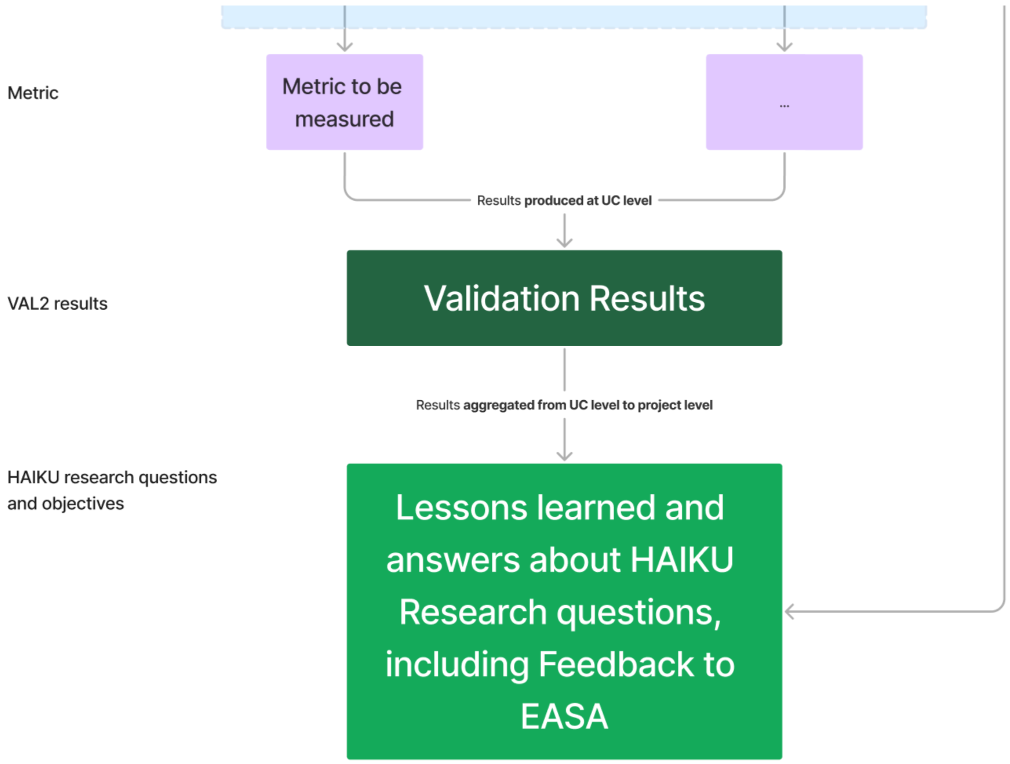


Table 1 provides an overview of the HAT concepts explored in VAL2, specifying the EASA HAT classification, communication and interaction method, explanation mechanism and use of AI.

Table 1. *Human AI teaming concepts explored in VAL2*

UC	EASA HAT classification	Communication method	Interaction method	Explanation mechanism	Use of AI
#1	2B: Collaboration	Bidirectional communication	Touch screen	Automatic activation of the FOCUS assistant based on physiological parameters. FOCUS adapts its action by monitoring the gaze behaviour.	Detection of startle and surprise

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	2A: Cooperative	Bidirectional communication	Touch screen	Operational Intentions and Key performance Indicators (KPIs)	On the Translators (human to machine and machine to human)
#3	2B: Collaboration	Bidirectional communication	Mouse & Keyboard, voice, vision	Storytelling explainer to explain the Digital Assistant for UAM Coordinator's (DUC) decision-making.	None, only conceptual.
#4	1B: Assistance	Bidirectional communication (ATCOs can change the sequences and the Intelligent Sequence Assistant - ISA - will communicate the changes accordingly. Viceversa, if ISA changes the sequence, this is also communicated to the ATCO).	Mouse & Keyboard	Explanations for sequence changes are provided to ACTOs at different levels of detail according to the Construal Level Theory (CLT) levels. The ATCO can access them on-demand by interacting with the electronic strips	Regression model for estimating aircraft arrival time.
#5	1A: Augmented	Dashboard visualisation and exploration by expert operational users	Visual displays (Dashboard suite of 4 displays), mouse, keyboard	Self-explanatory - DB designed with users. Manual being developed for new users.	Limited use at present, e.g. for feature extraction from incident reports.
#6	1B/2B: Assistance/Collaborative	Bidirectional communication between	Touch screen via the app	Routing sequence and likelihood of infection provision to passengers.	Classification of infection likelihood,

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

		COVAID and passengers with real time sensor information and explanations. Also a portal for health and safety managers	Location tracking	Q&As to engage passengers to gain trust with sensor information and other information regarding the decision process. Statistics, classification and XAI for health and safety operators.	routing recommendation, explainable AI (XAI) and other information available through AI chatbot.

To create a harmonised overview of how all use cases support and address human-AI teaming aspects in HAIKU, we have applied the EASA HAT requirements as a point of reference (see D4.1v2, the M24 update for reference). Table 2 overviews the Explainability and Human Factors requirements proposed by EASA machine learning guidance document (EASA, 2024).

Table 2a. *Generic EXP Requirements*

AI Level	EASA Obj	Generic REQT
1B-2B	EXP-10	System must provide adequate level of explanation for each enabled output
1B-2B	EXP-11	System must provide explanations in a clear and unambiguous form
1B-2B	EXP-12	System must provide explanations relevant to the assessment of the appropriateness regarding expected decision / action
1B-2B	EXP-13	System must be able provide explanations with abstraction level appropriate to task, situation, trust, and expertise of user
1B-2B	EXP-14	[if XAI functionality is adaptive/adaptable] System must provide means for the end user to customise explanation's level of abstraction

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

1B-2B	EXP-15	System must provide explanations in a timely manner according to situation, operator needs, and operational impact
1B-2B	EXP-16	System must provide explanations upon user request. When needed for understanding, the system must provide additional details on request.
1B-2B	EXP-17	System must provide valid explanations for each output relevant to the task
1A-2B	EXP-18	System must include training and instructions to handle indications of input/output monitoring
1A-2B	EXP-19	System must inform user of unsafe AI-based operating conditions, in a timely manner System must include training and instructions to recognize unsafe operating conditions. System must include training and instructions to address unsafe operating conditions.

Table 2a. *Generic HF Requirements*

AI Level	EASA Obj	Generic REQT
2A-2B	HF-01	System must be able to generate its own situation representation
2A-2B	HF-02	System must be able to reinforce the user's individual situation awareness
2B	HF-03	System must be able to support a shared situation awareness
2A-2B	HF-04	If the system makes a decision requiring validation based on procedures] System must obtain crosscheck validation from the user
2A-2B	HF-05	In NORMAL operations AND complex situations] System must be able to recognize when the user strategy is sub optimal, and propose an improved solution
2B	HF-06	In ABNORMAL operations AND complex situations] System must be able to identify the problem, share the (root cause) diagnosis resolution strategy, and anticipated operational consequences
2B	HF-07	System must be able to respond to poor decision-making by the user, by detecting, alerting and assisting the end-user

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

2B	HF-08	System must be able to propose alternative solutions and support its positions
2B	HF-09	System must permit short term modifications / acceptance of task allocation pattern
2A-2B	HF-10	If spoken natural language is used] System must be able to process and acknowledge user requests, responses, and reactions
2B	HF-11	If spoken natural language is used] System must be able to notify the user that he or she possibly misunderstood the information
2B	HF-12	If spoken natural language is used] System must be able to identify, via user actions, that there was a possible misinterpretation by the user
2B	HF-13	If spoken natural language is used] System must be able to resolve confirmed misunderstandings or misinterpretations of spoken natural language
2A-2B	HF-14	If spoken natural language is used] System must not interfere with user's other communications or activities
2B	HF-15	If spoken natural language is used] System must provide information regarding the associated AI-based system capabilities and limitations.
2A-2B	HF-16	If spoken natural language is used] System must use spoken procedural language syntax that can be easily learned and applied by the user
2A-2B	HF-17	If gesture language is used] System must use gesture language syntax intuitively associated with the command that it is supposed to trigger.
2A-2B	HF-18	If gesture language is used] System must be able to disregard non-intentional gestures.
2B	HF-19	If gesture language is used] System must be able to recognize end-user intention.
2B	HF-20	If gesture language is used] System must be able to acknowledge end-user intention with appropriate feedback.
2A-2B	HF-21	If spoken natural language is used] System must permit deactivation of spoken natural language modality
2B	HF-22	If spoken (natural or procedural) language is used] System must be able to assess the performance of the dialogue.
2B	HF-23	If spoken (natural or procedural) language is used] System must be able to transition between spoken natural language and spoken procedural language, depending on the performance of the dialogue, the context of the situation and the characteristics of the task.
2B	HF-24	System must be able to combine or adapt interaction modalities depending on task characteristics, operational events, and/or operational environment
2B	HF-25	System must be able to automatically adapt interaction modality to user states, situation context, and/or received user preferences

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

2A-2B	HF-26	System must minimise the likelihood of design-related user errors
2A-2B	HF-27	System must minimise the likelihood of HAIRM-related errors: - System Situation Representation must be congruent with User Situation Awareness - System must be tolerant to User executing System allocated tasks
2A-2B	HF-28	System must tolerate end user errors
2A-2B	HF-29	System must detect and tolerate end user errors
2A-2B	HF-30	System must, once an error is detected, provide efficient means to inform the end user.
2B	HF-31	System must be able to diagnose failures and present pertinent information to user
2B	HF-32	System must be able to propose a solution to detected failures
2B	HF-33	System must be able to support the end user in the implementation of the solution.
2B	HF-34	System must inform the user that system failure logs are maintained for subsequent analysis

1.2 EASA Validation Requirements

In Table 2 of D4.1 (HAIKU, 2024), an overview was provided of how the IA of the six HAIKU use cases address EASA objectives at a conceptual level.

In Table 3 below, we provide an overview of the high-level validation objectives, aligned with EASA HAT requirements, explored by the use cases in VAL2. The green highlighted boxes indicate which EASA objectives that respective use case is addressing in VAL2. As such, the green boxes also indicate to which EASA objectives use cases have answered “yes” to the “Covered in VAL2” question in the validation framework shown in Figure 1.

Table 3. *EASA HAT requirements coverage by Use Case in VAL2 validation objectives*

Obj.	Short description	UC1	UC2	UC3	UC4	UC5	UC6
EXP-10	Characterise explainability needs	✔	✔	✔	✔		✔
EXP-11	Clear and unambiguous presentation of explanations	✔	✔	✔	✔		✔

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

EXP-12	Demonstrate relevance of explanation for decision/action	✓	✓	✓	✓		✓
EXP-13	Define the level of abstraction of explanations according to task, situation, trust, expertise of user…	✓	✓	✓	✓		✓
EXP-14	Customisation of explanation level of abstraction (if XAI adaptability/adaptiveness is available)						✓
EXP-15	Define explanations timing according to situation, end user needs, operational impact	✓		✓	✓		✓
EXP-16	Enable explanation and details upon user request	✓	✓	✓	✓		✓
EXP-17	Ensure validity of explanation	✓	✓		✓		✓
EXP-18	Provide instructions/training to handle indications of input/output monitoring	✓				✓	
EXP-19	Provide timely information on unsafe operating conditions				✓	✓	✓
HF-01	IA situational awareness building	✓	✓		✓	✓	✓
HF-02	User situational awareness reinforcement	✓	✓	✓	✓	✓	✓
HF-03	Shared situational awareness building	✓		✓	✓	✓	✓
HF-04	Ability to submit decisions for cross-check validation		✓	✓		✓	
HF-05	Identify suboptimal strategy (normal operation) to propose/justify optimised solution				✓		
HF-06	Identify abnormal operation, share diagnosis, resolution strategy, anticipated consequences	✓				✓	✓
HF-07	Detect poor decision-making by the end user in a time-critical situation	✓				✓	
HF-08	Propose alternatives and support own positions	✓		✓	✓		
HF-09	Modify and accept modification of task allocation / task adjustments (instantaneous/short-term)			✓	✓		✓
HF-10	Provide indication of acknowledged user's intentions			✓			✓
HF-11	Notify possible misinterpretation from the end user						
HF-12	Detect misinterpretation from the end user, based on his/her responses or actions			✓			
HF-13	Resolve misunderstanding/misinterpretation			✓			
HF-14	Ability to not interfere in other communications or actions						
HF-15	Ability to provide information on AI-based system capabilities and limitations.						

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

HF-16	Syntax for spoken procedural language designed to ease end user learning	✓					
HF-17	Syntax for gesture language designed to be intuitive						
HF-18	Ability to disregard non-intentional gestures						
HF-19	Ability to recognise end-user intention when using gestures						
HF-20	Ability to acknowledge the end-user intention when using gestures						
HF-21	Enable spoken natural language deactivation to benefit other modalities						
HF-22	Ability to assess the performance of the dialogue						
HF-23	Contextually transition between spoken natural language and spoken procedural language						
HF-24	Combine or adapt the interaction modalities depending on task and operations			✓			
HF-25	Automatically adapt the modality of interactions to end-user states, preferences and situations			✓			
HF-26	Minimise the likelihood of design-related errors made by the end user	✓				✓	
HF-27	Minimise the likelihood of design-related errors related to HAIRM	✓				✓	
HF-28	Demonstrate tolerance to end user errors	✓		✓	✓	✓	✓
HF-29	Provide opportunities to detect end user errors				✓	✓	✓
HF-30	Provide efficient means to inform the end user about detected errors						
HF-31	Ability to diagnose failures and present the pertinent information to the end user				✓		
HF-32	Ability to propose a solution to the failure	✓					
HF-33	Ability to support solution implementation						
HF-34	Inform that logs of system failures are kept for subsequent 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

1.3 Validation plan overview

Table 4 provides a high-level overview on the VAL2 plans.

Table 4. *Use Case validation plan for VAL2*

UC	Segment	Stakeholders and nr of participants	Timeframe	Operational Phase	Type of scenarios explored (normal, non-normal)
#1	Airborne	6 to 10 airline pilots	2025-2030	During operations	Non-stabilized approach
#2	Airborne	6 to 10 airline pilots	2030-2050	During operations	Diversion due to airport runway closure
#3	UAM/U-space	6 to 10 air traffic controllers, 1-2 ATC traffic flow managers, 1-2 operators from relevant domains	2030-2050	During operations	Medical emergency requiring rerouting; vertiport on fire requiring a portion of the U-space to be closed; conformance issue with flight deviating from U-plan.
#4	Airport ATM	Air traffic controllers	2025-2030	During operations	Management of high traffic density and multiple sequence changes, with cases of pilots not complying with ATCO's instructions.
#5	Airport	Safety managers	2025-2030	Planning and post-operations	Incident types - taxiway selection errors, pushback errors, holdpoint busts
#6	Airport	Candidate Pilot acting as passengers and staff as health and safety managers	2025-2030	During operations	Crowded and overcrowded situations

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 Use Case #1 – Flight Deck Startle Response

In the cockpit, startling and surprising events can occur and trigger a "startle effect" among the crew. The startle effect can be defined as the first response to a sudden, intense stimulus. It triggers an involuntary physiological reflex, such as blinking of the eyes, an increased heart rate and an increased tension of the muscles (Koch, 1999). On the flight deck, the startle effect is often combined with a surprise that results from a disparity between a person's expectations and what is actually perceived (Horstmann, 2006). As the flight deck is the interface between highly automated complex systems and pilots, such disparity between the reality and crew members' expectations can have significant consequences on the safety of the flight. Startle and surprise reactions have played a key role in a significant number of accidents, including Loss-of-Control In-flight (BEA, 2012; KNKT, 2014).

Strategies have been put in place to minimise the consequences of a startling and surprising event. Pilots are made aware of, and trained for, this problem. Theoretical courses and simulator sessions on events that can trigger a startle effect are followed by student pilots. All through their career, line pilots also attend recurrent training sessions every 6 months. Most airlines identify three kinds of training to minimise the startle effect among pilots: (i) Crew resource management (CRM) to prevent surprise by building a good crew situation awareness through effective communication. (ii) Threat and error management (TEM), to prepare mitigation means in advance to react more easily to a startle. (iii) Basic flying skills and upset Prevention and Recovery training (UPRT) to recognize threatening aircraft state and recover from it with manual flying using core competencies.

Intelligent Assistants could play a key role in startle events by accompanying pilots in the application of their training, especially when pilots lose accurate situational awareness due to startle coupled with initial confusion over what is happening. The Intelligent Assistant could also act as an additional crew member in the case of single pilot operations and reduced crew situations, as well as occupying a key role in future Personal Planes and Sky Taxis, whose pilots will likely have far less training compared to current airline pilots.

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

2.1 UC#1 TRL Overview

Table 5. *UC1 TRL overview*

COMPONENT: Startle effect detection, Situation awareness (SA) augmentation, and Stress regulation support		
TRL	**Month**	**Activity to reach selected level**
1	1	Concept formulated at the project's start
2	6	Interviews and design meetings with pilots to produce prototype specifications
3	16	VAL1 with a first version of the prototype: • Situation Awareness Augmentation and Stress regulation support: test in an aircraft simulator involving 5 pilots • Startle detection module: laboratory test
4	28	VAL2 with a second version of the prototype: • Situation Awareness Augmentation and Stress regulation support: test in a realistic aircraft simulator with pilots in a relevant operational scenario • Startle detection module: artificial intelligence module trained with new datasets and tested with data collected in simulator in a relevant operational scenario
5	-	
6	-	

UC1 is targeting TRL4. The main factors inhibiting the achievement of TRL6 within the HAIKU project timeframe include the need for the Intelligent Assistant integration in a real aircraft to carry out flight tests, while adhering to development, cost, safety, and performance constraints inherent to the aerospace industry.

Specifically:

- Situation Awareness Augmentation and Stress Regulation support: more robust tests should be performed to assess the benefits of these functions in an actual aircraft.
- Startle Detection module: the available data are insufficient to achieve a more robust and accurate detection of the pilot's state.

2.2 UC#1 Validation 1 Results and Lessons Learned

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

Upon completion of VAL1, the main lessons learned that point to new design directions are reported in Table 6 below (from D6.2).

Table 6. *UC1 VAL1 lesson learned and new design directions*

Insight	Functional Requirements	Proposed solution for next iteration
The duration for cancelling FOCUS activation is too short.	Pilots should be able to cancel the assistant as it could create distraction if not needed.	Increase the delay for assistant activation and evaluate that it will not impact the assistant benefits.
The demand of resolving aircraft automation failures did not allow pilots to perform deep breathing and regulate their stress level.	Deep breathing is key to reach cardiac coherence and reduce stress.	Investigate novel ways to increase pilot's attention to breathing and explore new forms of breathing support to help pilots reduce their stress.
The visual cues displayed on the electronic flight bag were not seen by the pilot during manual piloting.	Visual cues can provide pilots with self-awareness about their stress level or information about the flight status. They may support better stress management and better situation awareness.	Change the display location and the design of such information according to the pilot's task.
The simulated tactile heartbeat feedback was welcomed by participants but the performance of such an approach was not investigated in VAL1.	Because of being inconspicuous and undemanding and easily deployable in the cockpit through existing technology, this may be one of the best solutions for regulating stress in the cockpit	Evaluate the impact of simulated tactile heartbeat feedback on stress regulation.
FOCUS provides too many eye-catching alerts on cockpit screens.	Provide timely information to pilots.	Implement a priority queue for visual alerts and combine visual and oral alerts.

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

2.3 UC#1 Updated Concept of Operations

Key R&D Needs and Objectives

Problem Statement - what is the problem to address with the Intelligent Assistant?

The Intelligent Assistant could help pilots to recover from startle and surprise effect proposing a collaborative way to mitigate consequences in several steps:

- A first level of assistance could consist in the digital assistant supporting the pilot in overcoming Startle and surprise thanks to an emotion regulation function. This function consists of Biofeedback techniques. Biofeedback is a mind-body technique that involves using visual or auditory feedback to teach people to recognize the physical signs and symptoms of stress and anxiety, such as increased heart rate, body temperature, and muscle tension. The assistant will guide the pilot to breath at a specific rhythm thanks to visual feedback and will passively reduce his/her stress thanks to a haptic wristband.
- To support the pilot in making sense of the situation after a startling and/or surprising stimulus, a second level of assistance could consist of maintaining and raising pilot Situation Awareness. Indeed, under startle and/or surprise, a loss of situation awareness is in some cases observed.

A related research question is to what extent should the startle assistant be explainable or provide an explainable layer? An explainable layer in the Human Machine Interface (HMI) of the assistant itself may help pilots regain the situational awareness, but it can also overload pilots with additional information.

Human Factors & Safety Risks – Potential risks and problems introduced by the Intelligent Assistant.

Trust: How could a pilot trust and make the appropriate action while being under the effect of a startling event?

Overload: in these situations, the assistant may end up increasing the workload, by providing an additional source of information.

R&D Objectives

UC1-OBJ-01/02/03/05 were addressed during VAL1 and will not be the main focus for VAL2. The list of all validation objectives is provided below, with VAL2 objectives highlighted in bold.

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 7. *UC1 Key R&D Objectives*

OBJ-ID	Validation Objective	Success Criteria ID	Success Criteria
UC1-OBJ-01	To assess the operational relevance of the solution from the CAT (Commercial Air Transport) pilots perspective in SPO (Single pilot Operations).	UC1-CTR-01	The solution is considered relevant by CAT pilots in SPO.
UC1-OBJ-02	To assess the acceptability of the solution from the CAT pilots perspective in SPO.	UC1-CTR-02	The solution is considered acceptable by CAT pilots in SPO.
UC1-OBJ-03	To assess the feasibility and integration of the solution in a relevant operational environment.	UC1-CTR-03	The solution is considered feasible and can be integrated in relevant operational environments.
UC1-OBJ-04 (VAL2)	To assess the effectiveness and efficiency of the assistant support in a relevant operational environment.	UC1-CTR-04	The solution is considered effective and efficient in supporting CAT pilots to overcome the startle and surprise effect in relevant operational environments.
UC1-OBJ-05	To assess the generalisation of the solution to multiple different scenarios.	UC1-CTR-05	The solution is considered to be useful in different operational conditions by CAT pilots.
UC1-OBJ-06 (VAL2)	To assess what is the best solution to reduce stress with a haptic feedback	UC1-CTR-06	An efficient solution is identified.
UC1-OBJ-07 (VAL2)	To evaluate the startle and surprise detection module	UC1-CTR-07	The startle and surprise detection module is able to detect and identify pilot's state in post-analysis of the simulator sessions.
UC1-OBJ-08 (VAL2)	To evaluate the explainability provided by the solution	UC1-CTR-08	The assistant's interface is understood by pilots

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 Concept Description

In the future of commercial aviation, new kinds of Operations and advanced cockpit designs will fundamentally change the flight environment. FOCUS will be integrated into this evolved framework to enhance pilot performance and safety. It will utilise artificial intelligence to monitor flight parameters and pilot state in real-time. By providing predictive analytics and timely interventions, FOCUS will assist pilots in overcoming startle and surprise effects, ensuring swift and appropriate responses to unexpected situations. As the role of the pilot gradually evolves into that of a mission manager, requiring rapid decision-making in emergency scenarios, FOCUS will provide crucial support. This assistant will be an integral part of the new cockpit design, equipped with intuitive HMIs and adaptive learning capabilities to support the solo pilot in complex and dynamic scenarios, ultimately enhancing situational awareness and decision-making in the next generation of commercial aviation.

Performance Targets

Table 8. *UC1 Performance Targets*

KPA	Category	KPI
Startle detection	MoE/MoP	Detection accuracy, detection rapidity
Surprise detection	MoE/MoP	Detection accuracy, detection rapidity
Pilot startle and/or surprise physiological recovery (partial incapacitation)	MoE/MoP	Recovery rapidity, recovery rate. Pilot acceptance, pilot performance on the operational task.
Pilot situation awareness sustainability/recovery	MoE/MoP	Subjective situation awareness assessment, pilot performance on the operational task. Rapidity to come back to a “normal” scan path. “Normal” scan path recovery rate.

High Level Validation Requirements

Only the High-Level Requirements that are specific to the VAL2 are listed here.

Table 9. *UC1-HLR-02*

HL-REQ-ID	UC1-HLR-02
Requirement	The pilot must have authority over the digital assistant at all times.

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

Rationale	To reduce impact of potential AI errors and to guarantee responsibility to the pilot.
KPA	Operational, Safety

Table 10. *UC1-HLR-03*

HL-REQ-ID	UC1-HLR-03
Requirement	The crew (AI and pilot) must be able to minimise the consequences of the startle and surprise effect.
Rationale	To handle an unexpected event in the cockpit, a key objective is to minimise the effect of startle and surprise for the pilot to be able to respond quickly and accordingly to the situation.
KPA	Safety, Human Performance

Table 11. *UC1-HLR-04*

HL-REQ-ID	UC1-HLR-04
Requirement	The pilot must be able to be assisted at different levels of support.
Rationale	To guarantee the right level of assistance and optimal HAT.
KPA	Safety, Human Performance

2.4 UC#1 Second Validation Plan (VAL2)

Validation Objectives

The following EASA HAT requirements will be addressed as VAL2 objectives.

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

EASA HF Requirements

Table 12. *UC1 Cooperation/collaboration capabilities (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
HF-01: FOCUS must maintain an updated record of environmental and user (group and individual) states, events, behaviours, and control loop histories	NO			
HF-02: FOCUS SA module must demonstrate SA modulation, by linking (eye gaze inferred) SA and aircraft flight state	YES	To evaluate the subjective effectiveness	Questionnaires	Qualitative and Likert scale
HF-03: FOCUS SA module must demonstrate SA modulation, by linking (eye gaze inferred) SA and aircraft flight state	YES	To evaluate the subjective effectiveness	Questionnaires	Qualitative and Likert scale
HF-06: FOCUS Detection module must inform user of the triggering logic (event or state based)	NO			
HF-06: FOCUS Detection module must inform user of corrective decision / action taken by the system	YES	To evaluate of the user saw the assistant activation and deactivation	Questionnaire	Qualitative
HF-07: FOCUS Detection module must respond on a startle-event timescale.	YES	To analyse the simulator session data	Post simulator session analysis	Accuracy and Latency of startle detection
HF-07: All three FOCUS modules must clearly signal their detect / alert/ assist functions to the user.	YES	To evaluate the usability of the assistant	Questionnaire	Qualitative and quantitative

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

HF-08: FOCUS SA module must demonstrate different (SA modulation) strategies, depending on changes in either eye point of gaze pattern, or aircraft state.	YES	To evaluate the subjective efficiency	Questionnaires	Qualitative and Likert scale

Table 13. *UC1 Procedural Language (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
HF-16: FOCUS SA module must use a limited and pre-defined syntax, to drive user's visual scan	YES	To evaluate the subjective effectiveness	Questionnaires	Qualitative and Likert scale

Table 14. *UC1 Error Management (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
HF-26: FOCUS Detection and SA modules must act in accordance with flightdeck design- and aircraft handling philosophies and standards.	YES	To evaluate the subjective effectiveness	Questionnaires	Qualitative and Likert scale
HF-27: FOCUS Detection and SA modules must act in accordance with flightdeck design- and aircraft handling philosophies and standards.	YES	To evaluate the subjective effectiveness	Questionnaires	Qualitative and Likert scale
HF-28: FOCUS Detection module must implement and update its control actions based on pilot error / non-compliance.	NO			

Table 15. *UC1 Failure Management (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric

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

HF-32: In case of Detection module failure, FOCUS alerting functionality (see HF-01 / 03 / 06 / 07) is tied to user override (deactivation) capability.	YES	To evaluate the usability of the system	Questionnaires	Qualitative and Likert scale
HF-34: FOCUS must maintain, on EFB, separate failure history logs for the Detection, Regulation, and SA modules.	NO			

XAI Requirements

Table 16. *UC1 Attributes of OpXAI (EASA Macroarea)*

EASA-like requirement (D4.1)	**Covered**	**Validation Objective**	**Activity**	**Metric**
EXP-10: FOCUS must be able to provide different levels of explainability, based on user needs, for the Startle / Incapacitation Detection module, Stress Regulation module, and Situation Awareness (SA) augmentation module	YES	To evaluate the usability of the system	Questionnaires	Qualitative and Likert scale
EXP-11: FOCUS explanations must be clear and understandable to the user, for the Detection module, Regulation module,. and SA module	YES	To evaluate the usability of the system	Questionnaires	Qualitative and Likert scale
EXP-12: FOCUS explanations must permit evaluation and assessment of system (Detection / Regulation / SA augmentation) actions	YES	To evaluate the usability of the system	Questionnaires	Qualitative and Likert scale
EXP-13: FOCUS must be capable of explaining its (Detection/Regulation/SA augmentation) actions at different levels of explainability, and with progressive levels of detail	YES	To evaluate the usability of the system	Questionnaires	Qualitative and Likert scale

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

EXP-15: FOCUS explanations must be presented with minimum delay, consistent with a startle event timeline.	YES	To evaluate the usability of the system	Questionnaires	Qualitative and Likert scale
EXP-16: FOCUS explanations, as presented on the EFB, must permit user-driven probing of (CLT level) explanation details	YES	To evaluate the usability of the system	Questionnaires	Qualitative and Likert scale
EXP-17: FOCUS must provide the user valid and reliable explanations, consistent with data inputs and outputs of the three modules (i.e. startle detection, stress regulation, and SA augmentation)	YES	To evaluate the usability of the system	Questionnaires	Qualitative and Likert scale
EXP-18: Training must include familiarisation with system I/O states, including correlation between input states and events (environmental startle events, pilot stress response) and system decisions and actions	YES	To tailor SA support to the pilot's need	Questionnaire	Ranking of the different SA support level

Validation Approach

VAL2 will consist of evaluating key components of the assistant during several activities:

Overall design validation: The assistant has been redesigned following VAL1 feedback from professional pilots. This new design will be tested by pilots in an A320 simulator. Operational relevance, integration in the simulator, acceptability, explainability of the interface and human-machine teaming will be evaluated through this activity

- **Stress regulation function validation**: An important function of the FOCUS assistant consists of supporting the pilot regulating his/her stress. During VAL1, a solution has been proposed to pilots but the objective was not to evaluate its efficiency. For VAL2, we will conduct a controlled experiment to assess if haptic feedback is an efficient way to reduce stress and to test several designs.
- **Startle and surprise detection validation:** During VAL1, performance of the startle detection module was not good enough to be implemented in the

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

cockpit. For VAL2, we will produce a specific startle and surprise dataset to train the startle detection module and test its performances on the data from the simulator sessions.

Exercise Description and Scope

Table 17. *UC1 Exercise Description*

Exercise Description	**Scope (What is being assessed)**	**Key R&D Objectives explored**
Experiencing startle and surprise in relevant operational environments with the assistant support	The usability, acceptability and integration of the assistant	UC1-OBJ-01/02/03/04
Testing several levels of support of the assistant	The usability, acceptability and integration of the assistant	UC1-OBJ-01/02/03/04
Testing the assistant's support in different operational conditions	The ability of the assistant to provide a relevant support in different flight phases	UC1-OBJ-05
Testing the HMI to provide an adapted level of explainability	The usability, acceptability and integration of the assistant	UC1-OBJ-01/02/03/04/08
Testing several designs of haptic feedback to reduce stress	The effectiveness of the assistant stress regulation function	UC1-OBJ-06
Producing a custom startle/surprise dataset, train the assistant and test the performances of the model	The effectiveness of the assistant in detecting startle and surprise	UC1-OBJ-07

Exercise Scenarios(s)

All simulator scenarios will be played on the ENAC A320 research simulator.

To evaluate the usability and the explainability of the assistant, three final approach scenarios without any startle/surprise stimuli will be played. Each approach will be performed with a different level of situation awareness support.

Participants will then choose their preferred level of support to perform a startling and surprising scenario.

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 chosen scenario is based on a scenario used in scientific literature to trigger startle and surprise effects. This scenario has been validated during VAL1 and is described below:

Lightning strike on final approach: On final approach, the aircraft is struck by lightning. As a result, a loud bang is heard, and an intense flash is triggered, provoking startle and surprise. Because of the lightning strike, electrical problems on board the aircraft lead to automatic disconnection.

Concerning the evaluations of the haptic feedback to reduce stress, the exact protocol is not defined yet.

Participants and their Role

For the simulator sessions, between 5 and 10 professional pilots will be recruited. Gender equality will be achieved if possible.

For the haptic feedback experiments, at least 20 people (not necessary pilots) will participate in the study.

Platform / Tools & Technique

The VAL2 simulator sessions will be performed in the ENAC A320 Research simulator.

Tobii glasses 2 will be used to track participants' gaze and a BITalino device will be used to gather participants' physiological data in all validation activities.

Intelligent Assistant

FOCUS is a flight deck system that will activate when a startle and/or a surprise is detected by pilots. Therefore, the assistant will be used in a degraded context, the aircraft being in an abnormal state and/or the pilot being partially incapacitated. If necessary and on demand, the FOCUS assistant can be activated/deactivated manually by the pilot.

Upon an unexpected event (e.g. lightning strike, system failures):

1. FOCUS detects the pilot's abnormal state.
2. The pilot maintains core task performance (FLY, NAVIGATE, COMMUNICATE).
3. In the meantime, FOCUS supports the pilot in regulating his/her stress by providing breathing guidance and haptic feedback.

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 Furthermore, FOCUS also helps in maintaining sufficient situation awareness by drawing attention towards unseen deviating flight parameters.
5 The collaboration between FOCUS and the pilots ends when the startle effect is over

Data Collection and Analysis Methods

The qualitative data collection will be done through debriefings held after each scenario and at the end of the session. The over-the-shoulder observations will be performed by Human Factors and Operational experts. The quantitative data will be collected through the platform's logs and recordings.

Observations: This technique mainly allows to address topics related to Human Performance, with the purpose to provide detailed and reliable information on the way the activity is carried out by the user. Direct observation enables gathering a high amount of data, especially qualitative data which cannot be collected through other methods.

In the validation exercises, direct over-the-shoulder observation will be used to collect insights about the pilot's performance, including aspects related to experienced workload, situation awareness, usability, faced difficulties, recovery actions, safety related events, etc.

Questionnaires (standard and ad hoc): After each scenario, pilots will be requested to fill in a questionnaire order to provide their feedback on aspects related to the assessment of Human-Machine Teaming, mental workload, situation awareness, trust, usability and safety.

The use of other dedicated (presumably custom designed) questionnaires might be considered to gather pilots' insight on other variables such as felt surprise or startle.

De-briefings: Debriefings, questionnaires and over-the-shoulder observations are interconnected techniques. This means that on the one hand, data collected through observations and questionnaires will be verified and discussed during the debriefings. On the other hand, insights extracted from the debriefings will be used to guide the following observations. This combination of techniques can complement and reinforce the quality of the quantitative data collected and contributes to achieving more reliable results.

Platform logs and participants physiological data: System quantitative data will be collected by the extraction of log files from the simulation platform. The data will be used as performance indicators and to further enrich the qualitative information collection (e.g. pilot input on the interface, reaction time, decision taken, etc.).

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

Moreover, participants' physiological data (e.g. heart rate, breathing rate, gaze behaviour, electro dermal response, muscles activity) will be gathered to assess the impact of the unexpected events and evaluate the support provided by the digital assistant.

Planned Activities

Table 18. *UC1 Planned Activities*

Activity	Activity description	General information
Overall design validation	Tests in the A320 simulator with several qualified pilots.	Toulouse, France, November 2024
Stress regulation function validation	Specific laboratories evaluation of several haptic feedback designs to passively reduce stress.	Toulouse, France, Q3 2024
Startle and surprise detection validation	Evaluation of the startle and surprise detection module fed with online data and ENAC and DFKI produced datasets	Q3-Q4 2024

Table 19. *UC1 GANTT*

ConOps & Requirement Definition
Design & Development
Validation

Year	2024						2025							
Month	7	8	9	10	11	12	1	2	3	4	5	6	7	8
Human-AI Teaming concept														
AI component/s														

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

Non-AI technological component/s															
HMI															

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 Use Case #2 – Flight Deck Route Planning/Replanning

With the constant growth in air traffic and climate change, pilots will face all sorts of new challenges. UC2 is based on a cockpit Intelligent Assistant (IA) to help the pilots in flight reroute and divert decision-making. The new routes will take into account numerous factors (e.g. remaining fuel available, distance to airport; in-route turbulence, connections possible for passengers given their ultimate destinations, airline hub, etc.). Here, the AI is used as an assistant for the pilots, so they remain in charge but communicate/negotiate with the IA to find the optimal solution according to their intentions. UC2 examines the potential benefits of integrating artificial intelligence (AI) and machine learning (ML) in aviation to enhance human autonomy teams (HAT). VAL1 enabled us to assess the effectiveness of varying levels of assistance (EASA AI/ML levels classification 1B, 2A and 2B) and interfaces.

3.1 UC#2 TRL Overview

Table 20. *UC2 TRL overview*

COMPONENT: Intelligent Assistant Concept		
TRL	**Month**	**Activity to reach selected level**
1	1	COMBI Technology component validated at TRL3 for Mission Management in Defense context with Human-in-the-loop simulations.
2	6	Concept definition of intelligent assistant for regional segment
3	16	Val1 with a exploratory study of 3 EASA Human-AI levels : Assistive, cooperative and collaborative
4	28	Val2 with the prototype for Human-AI cooperation in the diversion use case. - Test with 6-10 pilots
5	28	Val2 with the integration of the concept in a representative environment: Thales simulator using representative avionic systems: FlightX architecture and Flight Management System (FMS200). - Test with 6-10 pilots

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 concept of Intelligent Assistant based on COMBI is targeting TRL5. The proposed architecture is being deployed over a representative avionic scenario. Validation with 6-10 pilots will be performed by using the Thales flight simulator.

The main factors inhibiting the achievement of TRL6 within the HAIKU project timeframe is:

- Demonstration in an operational representative environment. For example, higher level of representativeness of interactions, namely integrating the operational control centre and ATC interactions.

3.2 UC#2 Validation 1 Results and Lessons Learned

The UC2-OBJ-02, OBJ-04 and OBJ-05 were addressed during the VAL1 experiment. The analysis permitted the identification of the key features associated with each type of assistance (1B decision support, 2A cooperative, and 2B collaborative) that facilitate the assurance and effectiveness of teamwork requirements. Pilots have indicated that they find all three types of assistance useful in general. However, they have preferences for specific features offered by different versions of these assistants. The pilots rated the interfaces as satisfactory, citing their ease of use, the effectiveness of interaction, and the accuracy of information provided. The table below evolves the insights from VAL1 (in D6.2) and those garnered during the development in WP4 (D4.3 and D4.5).

Table 21. *UC2 VAL1 lesson learned and new design directions*

Insight	**Functional Requirements**	**Proposed solution for next iteration**
Evaluation using operational intentions are very different between pilots (mean value as first hypothesis)	The system provides solutions that match with the mental model of pilots (according with the operational intentions)	Identify if a mean value of training data set is acceptable by the pilots during VAL2. Alternative is personalisation according to pilot preferences.

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 intentions are more impacted by trajectory	Evaluation of operational intentions are perceived by the pilot as a representative indicator for the global solution (selection of airport and trajectory)	Fine Tuning the model (COMBI translators) if needed according to the perception of the pilots.
The way to present complementary information should be improved.	Avoid information overload and enable rapid analysis and understanding to facilitate decision-making.	Conduct new interviews and workshop with pilots to determine the most concise and comprehensible way of presenting additional information on the interface
Pilots have appreciated the possibility to analyse and compare different solutions proposed by the IA.	Allow pilots to quickly compare multiple ranked solutions.	Present multiple solutions with a set of immediate information.
Various additional information was requested by the pilots.	Provide additional information needed to support pilot decision-making.	Investigate added value of customization of options presentation order, supporting different decision-making methods

3.3 UC#2 Updated Concept of Operations

The findings and insights of VAL1 will be used to develop a new interface, guided by pilots' feedback and Human Factors expertise. This interface may propose several solutions to the pilot that are easily and quickly comparable for the three intentions and the main associated KPIs. Additionally, the explanations of the AI with CLT levels will be further investigated, along with four levels of information representation:

1 Intentions/KPIs,
2 spatial,
3 temporal, and
4 resources

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

Key R&D Needs and Objectives

Problem Statement - what is the problem to address with the Intelligent Assistant?

Following the conclusion of VAL1, a data collection has been scheduled with 10 pilots. The objective of this session was to train the AI that will be developed and used for VAL2. Each pilot was required to assess 170 distinct images depicting a variety of flight scenarios, encompassing diverse routes and meteorological conditions. Furthermore, the participants were presented with a number of alternative airport solutions, each of which was evaluated in accordance with a number of key performance indicators (KPIs). For each image, participants were required to provide a value for each of the three intentions (Pilot cognitive comfort, Passengers comfort, Airline profitability), based on their expertise and experience. This new way to communicate with the IA via intentions with the COMBI (Bidirectional COMmunication) translator.

Human Factors & Safety Risks – Potential risks and problems introduced by the Intelligent Assistant

In document D6.1, two potential risks and problems introduced by the intelligent assistant were identified:

- Defining the Intelligent Assistant usage envelope, according to the AI/ML limitations with regard to situation variability.
- Use Experience of the Intelligent Assistant communication at the level of intentions. - Acceptability by pilots, including development of trust.
    - o The training of AI systems with 10 pilots of different nationalities, airlines and experience levels should provide training data that is sufficiently representative of the pilots in question. The acceptability and trust in this trained IA will be evaluated in VAL2.

R&D Objectives

The UC2-OBJ-02, OBJ-04 and OBJ-05 were the subjects of investigation during the VAL1 experiment. During VAL2, the OBJ-01 and OBJ-03 will be addressed in accordance with the specifications outlined in D6.1.

Table 22. *UC2 Key R&D Objectives*

OBJ-ID	Validation Objective	Success Criteria ID	Success Criteria
UC2-OBJ-	Combi enable	UC2-CRT-DT	Comparison against benchmark

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

01	effective and efficient high-level intentions communication in the team.	(Decision Time) UC2-CRT-RT (Reaction Time) UC2-CRT-PWL (Perceived Workload) UC2-CRT-UAC C (User Acceptance) UC2-CRT-RAC C (Rate of Acceptance)	(e.g., SPO or 2P crew direct inputting technical parameters into solver without assistance). User Acceptance: Questionnaire validation (SUS, CSUQ, TAM3 ...) [Rate of Acceptance: Acceptance and rejection rates for AI recommendations by the pilots (trust and why they do not use it (a way to improve HAT)].
UC2-OBJ-03	Methods: HAT design methodology able to support HAT safety and effectivity assessments	UC2-CTR-03	Comparison against benchmark (SPO or 2P crew) and decision support assistance

The validation plan described in section 5.3 and following focuses on UC2-OBJ-01.

Regarding the UC2-OBJ-03, an evaluation with experts, regarding the HAT design methodology will be pursued. Namely the following items will be evaluated:

- Does the methodology produce design information regarding HAT safety and effectiveness of HAT? (i.e., does it enable engineering of HAT from safety and effectiveness perspectives?)
- Does the methodology contribute to the requirements definition?
- Does the methodology enable evaluation and guidance of HAT architectural design for HAT? (e.g, does methodology support HAT allocation decisions?)

Operational Concept Description

The operational context employed in VAL1 will be replicated in VAL2. The flight in question will be a regional European flight to an airport that experiences a number of challenging meteorological conditions throughout the year. Due to inclement weather, the pilots are compelled to deviate from the originally planned route during the cruise. As the flight approaches its destination, the weather deteriorates, necessitating a diversion to another airport. Pilots will be presented with these two events, which will require the assistance of the IA. In one condition (with COMBI), the IA will communicate via intentions, whereas in the absence of COMBI, it will do so via more technical terms. In each condition, the IA will propose a new route and a new alternate airport, based on the pilots' preferences/needs/intentions.

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 experiment will be conducted in a flight simulator provided by Thales, designated FlytX. This simulator simulates an A320 with a wide-screen avionics suite. The IA will be able to upload the new flight plan directly into the FMS via ARINC compliance link, without the need for manual inputs from the pilots.

Performance Targets

In D6.1, a number of Key Performance Areas (KPAs) were identified, pertaining to both Measures of Effectiveness (MoEs) and Measures of Performance (MoPs) of the HAT concept. They remain unaltered from their state as presented in D6.1, and have been completed with the identification of certain KPIs during the pilot interviews conducted for VAL1.

Table 23. *UC2 Key Performance Areas*

KPA	Category	KPI
Mission Safety	MoE	Safety margin against State-of-Practice (SME evaluation)
Mission Performance	MoE	Operational impact index [function of cost efficiency, punctuality, passenger experience, accommodations, airline hub, distance from original destination, etc...](SME evaluation)
Decision Quality	MoP	Decision quality index (function of aspects taken into account, as evaluated by SME) Implementation Feasibility index (function of time to decide, time to implement, ATM considerations, airline consideration, objectives compliance)(data from experiments)
Regulatory Acceptance	MoE	Beyond 2030: Self evaluation against the SoA, and an acceptance of SMEs from regulatory organisations.
Social Acceptance	MoE	Acceptance of the assistance due to perceived usefulness (reliability, trust, performance, etc, as evaluated by SME)

High Level Validation Requirements

The high level requirements for VAL2 are:

Table 24. *HAIKU_UC2_HLR_1*

HL-REQ-ID	HAIKU_UC2_HLR_1
Requirement	The crew must achieve a safe flight termination.

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

Rationale	Flight safety is not negotiable. The selected flight plan must be constrained by acceptable bounds of safety-related KPIs (e.g. final reserve fuel).
KPA	Mission Safety

Table 25. *HAIKU_UC2_HLR_2*

HL-REQ-ID	HAIKU_UC2_HLR_2
Requirement	The crew must minimise the impact of flight plan changes in company operation.
Rationale	The overall objective is to maximise the company goals under constrained contexts.
KPA	Mission Commercial Performance

Table 26. *HAIKU_UC2_HLR_3*

HL-REQ-ID	HAIKU_UC2_HLR_3
Requirement	The crew decision quality must be better than that of a crew with no assistance in the same situation.
Rationale	The team decision must be best-in-class. Different companies define different operational goals and their prioritisation. The team solution purpose is to achieve a good solution within the constraints imposed by the specific context where the decision must be made.
KPA	Decision Quality

Table 27. *HAIKU_UC2_HLR_4*

HL-REQ-ID	HAIKU_UC2_HLR_4
Requirement	The crew decision must be implementable, considering time constraints.
Rationale	The decision must be provided with enough time to enable implementation (by human, by AI) after stakeholders validation (ATM and/or AOCC), when needed.
KPA	Decision Quality

Table 28. *HAIKU_UC2_HLR_5*

HL-REQ-ID	HAIKU_UC2_HLR_5
Requirement	The operational concept must be compliant to regulations applicable beyond 2030

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

Rationale	New regulations and/or guidance, namely on AI-based applications, must be used as reference for Systems Requirements.
KPA	Social Acceptance

3.4 UC#2 Second Validation Plan (VAL2)

Validation Objectives

The EASA-like requirements defined in D4.1 (HAIKU, 2024) to be validated in this phase are detailed below.

EASA HF Requirements

Table 29. *UC2 Cooperation/collaboration capabilities (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
HF-01: IA must account for weather impact upon flight trajectory.	YES	HAIKU_UC2_HAT_7: The IA solution must account for weather and destination states' impact in diversion and flight trajectory proposals.	Pilot assessment of solutions consistency in reference situations.	SME interview (self-confront ation with pilots)
HF-01: IA must account for pilot operational intentions.	YES	HAIKU_UC2_HAT_2: The IA and the Pilot must be able to communicate bi-directionally through operational intentions.	Pilots usability evaluation regarding intentions communication	Use a standard questionnaire: • CSUQ • SUS • UMUX
HF-01: IA must account for pilot operational intentions	YES	HAIKU_UC2_HAT_9: The IA interface must allow the pilot to recalibrate operational intentions during operation.	Requirement validation: Pilots usability evaluation; Requirement verification: System behaviour inspection	Number and type of 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

HF-02: IA must incorporate bidirectional information sharing (to/from the PF and PM (if appropriate)) about reroute / alternate airport recommendations, so as to match (pre-flight loaded) operational intentions and technical parameters.	YES	HAIKU_UC2_HAT_2: The IA and the Pilot must be able to communicate bi-directionally through operational intentions.	Pilots usability evaluation regarding intentions communication	Use a standard questionnaire : • CSUQ • SUS • UMUX
HF-04: IA recommendations must be approved by the user	YES	HAIKU_UC2_HAT_11: The IA must be able to load a Flight Plan approved by the Pilot to the FMS.	Pilot evaluation of explanations relevance, clarity and validity; Pilots trust evaluation	SME interview (self-confront ation with pilots) "Trust in IA" questionnaire

EASA XAI Requirements

Table 30. *UC2 Attributes of OpXAI (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
EXP-10: IA must be able to explain both its reroute- and airport diversion recommendations using progressive CLT levels.	YES	HAIKU_UC2_HAT_4: The IA must provide relevant, clear, and valid explanations to the Pilot about each proposed solution.	Pilot evaluation of explanations relevance, clarity and validity; Pilots trust evaluation	Gap between expected pilot's mental representation
EXP-11: IA must provide explanations in a clear and unambiguous form	YES	HAIKU_UC2_HAT_4: The IA must provide relevant, clear, and valid explanations to the Pilot about each proposed solution.	Pilot evaluation of explanations, clarity	Likert scale
EXP-12: IA must provide explanations relevant to the assessment of the appropriateness regarding expected decision / action	YES	HAIKU_UC2_HAT_4: The IA must provide relevant, clear, and valid explanations to the Pilot about each proposed solution.	PIlot evaluation of explanations relevance; Pilot trust evaluation	Likert scale

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

EXP-13: IA must provide explanations at the level of operational intentions along with each proposed solution.	YES	HAIKU_UC2_HAT_4: The IA must provide relevant, clear, and valid explanations to the Pilot about each proposed solution.	Observation	The interface displays the explanations for each solution.
EXP-14: Customisation of explanation level of abstraction (if XAI adaptability is available)	NO	Customization capability is not available in concept		
EXP-15: Define explanations timing according to situation, end user needs, operational impact	NO	Scripted situation, impossible to assess the needs of the situation, nothing to assess the needs of the users.		
EXP-16: IA interface must provide the Pilot the ability to navigate, for each proposed solution, through three different layers of explanations.	YES	HAIKU_UC2_HAT_6: The IA interface must provide the Pilot the ability to navigate, for each proposed solution, through different layers of explanations.	Observation	OK/NOT_OK
EXP-17: IA must provide timely, relevant and valid explanations to the Pilot about each proposed solution.	YES	HAIKU_UC2_HAT_4: The IA must provide relevant, clear, and valid explanations to the Pilot about each proposed solution.	Observation	Likert Scale
EXP-18: The pilot must demonstrate adequate knowledge level on handling IA monitoring on ODD and confidence levels.	NO	Not addressed in VAL2		
EXP-19: IA must provide indication that it is operating unsafely.	NO	No unsafe situations will be scripted into the validation		
EXP-19: IA training must cover recognition and handling of unsafe operating conditions.	NO	No unsafe situations will be scripted into the validation		

Non-EASA like requirements

In addition to the EASA-like requirements, the following requirements have been defined for UC2 that will be explored in VAL2.

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 31. *UC2 Non-EASA Like requirements HAIKU_UC2_HAT_1*

HL-REQ-ID	HAIKU_UC2_HAT_1
Requirement	When requested or after identification of diversion and/or re-route, the IA must provide three solutions according to the operational intentions selected by the Pilot.
Rationale	Supports TEAMWORK performance; Calculation abilities to provide the decision support
Verification Method	Inspection, High Fidelity Human-In-The-Loop (HITL)
Verification Criteria	SME assessment of solutions consistency in reference situations.
HLR	HAIKU_UC2_HLR_3, HAIKU_UC2_HLR_4

Table 32. *UC2 Non-EASA Like requirements HAIKU_UC2_HAT_5*

HL-REQ-ID	HAIKU_UC2_HAT_5
Requirement	The pilot must demonstrate adequate knowledge level on the IA capabilities and typical behaviour.
Rationale	Baseline understanding of the assistant capabilities, limitations and expected behaviour must be built to enable effective and efficient decision-making.
Verification Method	High Fidelity HITL
Verification Criteria	Observation of behaviour during experiment; SME evaluation on pilot ability to interact with the IA and properly use the IA.
HLR/EASA HLR	HAIKU_UC2_HLR_2, HAIKU_UC2_HLR_3, HAIKU_UC2_HLR_4

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

Validation Approach

The IA will be integrated in a highly representative simulator provided by Thales (FlyX). This simulator will provide a simulation environment that integrates real systems of a current cockpit. The bidirectional communication concept (COMBI) will be implemented in a specific user interface based on the results of VAL1, with the objective of evaluating the global concept of HAT based on operational intentions.

Exercise Description and Scope

The high fidelity HAT for VAL2 will be tested.

Table 33. *UC2 Exercise Description*

Exercise Description	Scope (What is being assessed)	Key R&D Objectives explored
Experience a diversion	Decision-making, usability, acceptability, trust	UC2-OBJ-03
Testing bidirectional communication via intentions and via main KPIs	Communication effectiveness, efficiency, usability, acceptability	UC2-OBJ-01
Testing OpXAI implementation	Explainability need, level of CLTs required, trust	UC2-OBJ-03

Exercise Scenarios(s)

The flight scenario will be done on the Thales flight simulator. The flight scenario will be the same as in VAL1, except that the departure and arrival airports may be changed. However, the events will be equivalent to VAL1 as presented in detail in D6.1.

Participants and their Role

Ten pilots should participate in the VAL2. Each of them will perform several similar flight scenarios with and without COMBI to compare the situations.

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

Platform / Tools & Technique

The Thales simulator will be configured to run the UC2 scenarios. This avionics platform can integrate the developed concepts in order to test it with Operational Expert.

This modular platform will integrate very representative components of avionics, that are today "real products". One example of that is the Thales FMS200.

Our ambition is to demonstrate that the proposed concept of Intelligent Assistant in cockpit operations has a high-maturity level to be integrated in a realistic avionic system (Thales FlightX).

Figure 2. *Thales FlightX*

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

Intelligent Assistant

The IA will provide assistance to the pilot on two separate occasions. Initially, the pilot will request a new route to circumvent adverse weather conditions along the initial flight plan route. In the baseline scenario, this request will be made in accordance with a set of pre-defined KPIs. In the COMBI condition, the pilot will initiate the request by selecting the appropriate intentions. In the second phase of the scenario, the pilot is required to divert the aircraft and select an alternative airport. Once more, in the baseline condition, the pilots must assess potential solutions through the application of KPIs. In contrast, in the COMBI condition, they must evaluate the proposed outcomes of the IA, as indicated by the intentions.

Data Collection and Analysis Methods

As for VAL1, qualitative data will be gathered directly after the simulator sessions, during walkthroughs by human factors and operational experts. Quantitative data will be gathered via: logs of user interactions during the experiments and questionnaires following the completion of the walkthroughs. The data will be subjected to statistical analysis to compare the IA with and without high level intention communication. The questionnaires will evaluate the usefulness, usability and acceptability of the IA.

Planned Activities

Table 34. *UC2 Planned Activities*

Activity	Activity description	General information
Finalise IA development	Fine-tuning of the AI and validation tests	Bordeaux, France, Q3 2024
Development of the VAL2 HMI	Workshop of design thinking for the HMI design and development	Bordeaux, France, Q4 2024
Validation of the experimental set-up	Technical validation of the test bench, checking communication between components, information gathering and a full dry run test.	Bordeaux, France, Q4 2024
VAL2	The participants will engage in the VAL2 experiment within the flight simulator, with the objective of testing the IA.	Bordeaux, France, Q1 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

Table 35. *UC2 GANTT*

Legend
ConOps & Requirement Definition
Design & Development
Validation

Year	**2024**						**2025**							
Month	7	8	9	10	11	12	1	2	3	4	5	6	7	8
Human-AI Teaming concept														
AI component/s	Validation	Validation	Design & Development	Design & Development	Design & Development	Validation	Validation							
Non-AI technological component/s	Validation	Validation	Validation	Design & Development	Design & Development	Validation	Validation							
HMI	Validation	Validation	Validation	Design & Development	Design & Development	Validation	Validation							

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 Use Case #3 – Urban Air Mobility

Drones and vertical take-off and landing (VTOL) aircraft are envisioned to drastically change how people and goods move in cities. Airborne transportation means in urban areas are expected to increase dramatically within the next decades, necessitating novel forms of traffic regulation. This new air transportation system, Advanced Air Mobility (AAM) with its subcomponents of Urban Air Mobility (UAM) and Unmanned Aircraft Systems (UAS) is expected to go live on a smaller scale in cities during the late 2020s, initially manned by pilots and then as autonomous and unmanned from 2030. Aircraft are foreseen to be autonomous with little need for conventional air traffic management services such as conflict detection and resolution. Therefore most airspace services are assumed to be automated and require little human intervention.

Certain services, however, may necessitate greater human involvement due to the unpredictable and dynamic nature of the operations. In UC3 we embrace a research-through-design approach and Human Factors perspective to explore human-intelligent assistant teaming in urban airspace traffic management (i.e., U-space), starting during the transition period from the current state, where UAM and UAS operations are absent, to the envisioned bustling ecosystem by the mid or latter part of the 21st century. We envision a new human role - the UAM Coordinator who will collaborate with the intelligent assistant DUC (Digital assistant for UAM Coordinator). The UC3 U-space ConOps is based on the European CORUS-XUAM U-space ConOps.

4.1 UC#3 TRL Overview

Table 36. *UC3 TRL Overview*

COMPONENT: Traffic situation display (UTM City), Storytelling explainer system, DUC HAT HMI.		
TRL	**Month**	**Activity to reach selected level**
1	12	Literature review of UAM/UAS concept of operations, interviews with external stakeholders (EVE Air mobility) to identify futuristic scenarios, participatory design approach (workshops, interviews, and design meetings) involving domain experts (ATM, ATC, UAS, UAM, UTM) to define intelligent assistant concept and futuristic traffic scenarios.
2	18	Participatory design approach (workshops, interviews, and design meetings) involving domain experts (ATM, ATC, UAS, UAM, UTM) to determine and design prototype HMI specifications, including Storytelling explainer, and design prototype working environment.

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	24	VAI1 with a proof of concept for the IA prototype HMI and UAM Coordinator working position. Laboratory tests with air traffic controllers. Basic IA HAT HMI functionalities. IA backend was scripted and *Wizard of Oz* simulated. Workshops and interviews with key stakeholders (U-space developers, emergency responders, UAM manufactures) to iterate on concept development, in particular IA HAT HMI specifications, and the traffic situation display. Hazard and Operability Study (HAZOP) to further define IA HAT HMI requirements.
4	31	VAL2 with a second version of the IA prototype and UAM Coordinator working position. Traffic situation display, IA HAT HMI, and Storytelling explainer: test in UAM Coordinator working position simulator using relevant operational scenarios.
5	-	
6	-	

COMPONENT: DUC backend (AI components)		
TRL	**Month**	**Activity to reach selected level**
1	31	Literature review of UAM/UAS concept of operations, interviews with external stakeholders (EVE Air mobility) to identify futuristic scenarios, participatory design approach (workshops, interviews, and design meetings) involving domain experts (ATM, ATC, UAS, UAM, UTM) to define intelligent assistant concept and futuristic traffic scenarios.
2	-	
3	-	
4	-	
5	-	
6	-	

For VAL2, UC3 is targeting TRL4 concerning the components: Traffic situation display (UTM City), Storytelling explainer system, and the IA HAT HMI. The IA AI components will not be included in VAL2, but implemented through a wizard of Oz approach.

The main factors inhibiting the achievement of TRL6 within the HAIKU project timeframe 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

- Futuristic roles and working environment: As the envisioned working environment and human roles do not yet exist, their concepts must be developed alongside the IA. With no current U-space airspaces or active U-space users (i.e., traffic), a reference U-space environment and traffic scenarios need to be defined to establish relevant reference scenarios.Air traffic controllers, flow managers, and other traffic management operators serve as proxy end users in validation activities, standing in for the currently non-existent UAM Coordinators.
- Concept maturity: The IA concept's broad scope makes it difficult to define all its aspects comprehensively, and determine which aspects to focus on in prototype development. Additionally, the target working environment and human role are not yet established, leading to unclear objectives and requirements for the system's functions or the problems it is intended to address.
- Resource constraints: Limited time and funding restrict the in-depth exploration and development of a fully mature prototype, especially given the broad scope of the IA concept. Resource constraints also limit the exploration and development of AI components.

4.2 UC#3 Validation 1 Results and Lessons Learned

The table below outlines the main lessons learned and new design directions derived from VAL1 (also in D6.2).

Table 37. *UC3 VAL1 lesson learned and new design directions*

Insight	Functional Requirements	Proposed solution for next iteration
Participants wanted to know more about what DUC was doing (actions taken)	DUC must provide information on what it is doing	While it would be inappropriate to provide information on all DUCs actions because of the large number of actions conducted in parallel, DUC could provide high-level information on which tasks, services, and goals are currently worked upon. This function should be on-demand

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

Participants wanted more information on DUCs' reasoning underlying recommendations.	DUC must be able to present which factors are considered, and how they are weighted to derive a particular recommendation.	Consider adding a more detailed feature list of factors considered by DUC, in order of importance. This should also be a function on request.
Participants requested the ability to access more detailed information about UAM HMI constructs, such as aircraft, vertiports, geofences, and hospitals.	Iterate on scenario design to provide more information about relevant UAM related constructs that can impact UAM Coordinators decision making.	Integrate more detailed information about UAM constructs in HMI. Should be accessible on demand.
DUC's voice and pop-up dialogue windows were appreciated by participants, but at times they appeared to miss a message from DUC (either not hearing it or not looking at the pop-up window).	To achieve closed loop communication, DUC must be able to verify that the human has acknowledged/received the information provided.	Implement a function for the human to verify messages from DUC. This can be achieved by "check" buttons, a verbal repeat/check function, and supported by DUC monitoring the human's attention through the use of eye tracking.

4.3 UC#3 Updated Concept of Operations

We have applied the CORUS-XUAM concept of operations to design a U-Space for Stockholm, Sweden, projected for 2030-2050. To ensure safety for U-space users and city residents, we propose a centralised safety and coordination function, termed the U-space Control Center. This centre, operated by the U-Space Service Provider (USSP), manages the airspace and delivers U-space services. The UAM Coordinator, supported by the intelligent assistant DUC, ensures these services. Their collaboration follows EASA's human-AI teaming level 2B. Together, the UAM Coordinator and DUC function as a team to achieve the U-Space goals:

- Ensure the safety of the U-space, its users, and affected individuals (e.g., citizens) by preventing accidents, managing risks, and adhering to emergency procedures, achieved through the provision of various U-space services (see section 6.2.2).

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

- Ensure efficient U-space utilisation by optimising its usage to maximise capacity while minimising delays and environmental impact.
- Enhance U-space security by monitoring for unauthorised access, detecting malicious activities, and identifying potential threats.

The revised ConOps includes updates to both the DUC and the UAM Coordinator roles. The shared goals and high-level tasks for their collaboration have been more clearly defined. Two main categories of high-level tasks were identified:

1. **U-space services (traffic management):** These tasks are crucial for achieving the overarching goals and include dynamic capacity management, emergency management, traffic information, and conformance monitoring.

2. **Team Collaboration Requirements:** This category focuses on establishing effective human-AI teaming between the DUC and the UAM Coordinator, building on situation awareness, transparency, decision-making, and bi-directional communication. For instance, eye-tracking technology will be employed in VAL2 to assist the DUC in enhancing the UAM Coordinator's situational awareness and establishing a shared situational understanding through attention-guidance mechanisms.

Additionally, the updated ConOps introduces a UAM Coordinator workstation featuring three screens: a knowledge/checklist library, a situation display based on UTM City, and a logbook with storytelling explanations.

Key R&D Needs and Objectives

The main R&D objective of UC3 is to explore through design an operational concept that enables effective human-AI teaming and collaboration, integrating human operators with intelligent assistance to ensure the viability of critical UAM operations within U-space traffic management. This concept explores how human-AI collaboration can be achieved, using representative scenarios that depict key U-space operations and employing a prototype intelligent assistant to simulate future system capabilities. The focus is on the U-space traffic manager role, embodied by the UAM Coordinator, with research efforts concentrating on designing the human-AI teaming interface to facilitate seamless interaction, shared situation awareness, and collaborative decision-making.

Problem Statement – what is the problem to address with the Digital Assistant?

Key challenges in achieving effective human-IA teamwork involve establishing and maintaining shared situational awareness, ensuring the UAM Coordinator's understanding of the DUC's actions and reasoning (i.e., transparency), enabling clear

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 effective communication, and fostering collaborative problem-solving. We envision that the UAM Coordinator will operate at a higher level of abstraction, such as allocating tasks and adjusting task thresholds and constraints for DUC. But it also involves working together with DUC on checklists to address emergency situations. DUC should be capable of directing the UAM Coordinator's attention to critical events or overlooked information. Generally, the UAM Coordinator's role would not involve detailed monitoring or direct interaction with individual flights.

Human Factors & Safety Risks – Potential risks and problems introduced by Intelligent agents in UAM.

- **Defining the operational boundaries of the intelligent assistant**: Establishing mechanisms to identify when the IA encounters situations beyond its capabilities, necessitating human intervention, particularly in cases of novel or unforeseen scenarios.
- **Identifying human communication requirements**: Effective human-IA collaboration relies on clear, bidirectional communication to prevent misunderstandings. How should the IA's interface be designed to support such communication, and what interaction methods should be implemented?
- **Ensuring human comprehension of the intelligent assistant's behaviour and reasoning**: Providing transparency/explainability to prevent confusion, misunderstandings, and unexpected automation outcomes.
- **Addressing over-reliance or mistrust of AI**: Balancing trust to avoid excessive dependence or mistrust toward the IA.
- **Managing operator workload**: Ensuring the human operator's workload remains manageable, especially during unexpected control handovers.
- **Maintaining situational awareness**: Strategies to keep human operators informed about the operational environment and IA actions, while also fostering share situation awareness between the human and IA.
- **Clarifying task allocation**: Defining roles, responsibilities, and task delegation to structure teamwork between the human and the IA.

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

R&D Objectives

Table 38. *UC3 Key R&D Objectives*

OBJ-ID	Validation Objective	Success Criteria ID	Success Criteria
UC3-OBJ-01	To assess the operational feasibility and acceptance of the IA concept.	UC3-CTR-01	Subjective feedback from participants acting as the UAM Coordinator with respect to supervising DUC and DUC's ability to support and establish a shared situation awareness, support decision making, afford transparency, and engage in bi-directional communication. Subjective feedback from participants acting as the UAM Coordinator with respect to DUC's ability to conduct high-level tasks related to monitoring, managing, and providing information to traffic in the U-space in an acceptable way.
UC3-OBJ-02	To assess the tasks and operating methods of the UAM Coordinator.	UC3-CTR-02	Subjective feedback from participants acting as the UAM Coordinator that the UAM Coordinator working position and content of the operating methods has been determined to be clear and consistent by domain experts and validation participants.
UC3-OBJ-03	To assess the UAM Coordinator timeliness of actions, workload, situational awareness, trust and acceptance.	UC3-CTR-03	Subjective feedback from participants acting as the UAM Coordinator that tasks can be performed, and problems solved, in an accurate, efficient and timely manner. Positive feedback from participants acting as the UAM Coordinator that: • workload is at an acceptable level. • situational awareness is sustained.
UC3-OBJ-04	To assess the effectiveness of the DUC in supporting the UAM Coordinator in managing emergencies	UC3-CTR-04	Subjective feedback from participants acting as the UAM Coordinator that DUC provides adequate support to handle the emergency without compromising safety. For rerouting recommendations, DUC should be able to provide information about the most suitable route, taking into account traffic, ground activity/availability, distance and airspace restrictions, aircraft state/capability (battery endurance).

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-OBJ-05	To assess the DUC HMI interface and information requirements.	UC3-CTR-05	Subjective feedback from participants acting as the UAM Coordinator that the IA HMI interface is user-friendly and provides necessary information.

Operational Concept Description

The UAM Coordinator and DUC provide high-level tasks in terms of U-Space services. The high-level tasks that the UAM Coordinator and DUC work with are defined by the U-space services to be provided to U-space users:

- Network identification (i.e., surveillance) of U-space users
- Tracking aircraft
- Geo-awareness management
- Dynamic capacity management
- Vertiport availability information
- Tactical conflict prediction/detection
- Tactical conflict resolution
- Emergency management
- Conformance monitoring
- Traffic information
- Vertical information and alerting
- Navigation infrastructure monitoring
- Legal monitoring and recording
    - Air status information
    - Digital logbook
    - Incident/accident reporting

In addition, there are information provision services provided by other stakeholders that the DUC and UAM Coordinator retrieve information from. The information gathered includes:

- Registration
- Drone aeronautical information management
- Common information service provider (CISP)
- Weather information
- Geospatial information
- Population density map
- Electromagnetic interference information

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

We anticipate varying levels of engagement from the DUC and UAM Coordinator in accomplishing above tasks. Generally, DUC is expected to autonomously monitor airspace conditions, track U-space users (including UAS and UAM flights), and provide real-time updates on airspace restrictions and capacity metrics. The UAM Coordinator will oversee DUC's actions, intervening when necessary and setting thresholds for when DUC can operate autonomously, especially in conflict prediction and resolution. Certain high-level tasks, such as managing conformance issues and dynamic geo-awareness, will require more collaboration between DUC and the UAM Coordinator to solve. For instance, DUC will monitor flight adherence to authorised U-plans, alerting both operators and the UAM Coordinator to any deviations. Additionally, DUC will track scheduled changes to geo-zones and no-fly areas, while enabling the UAM Coordinator to implement necessary constraints on short notice if needed. In dynamic capacity management, DUC will visualise airspace capacity and demand, making recommendations for adjustments based on various factors, while the UAM Coordinator will set constraints and determine suitable solutions for any necessary changes.

Another set of high-level tasks that will engage both the UAM Coordinator and DUC involves legal recording tasks, which focus on documenting events in the U-space. This data will be logged in a digital logbook and air status reports, with incidents and accidents reported separately. While DUC will primarily oversee and distribute these recordings, the UAM Coordinator must be able to supplement the information in real-time or afterwards, adding context to DUC's entries. The UAM Coordinator will use the Air Status Report to initiate changes in U-space operations, such as adjusting capacity or modifying constraints. The digital logbook, accessible on a separate screen, will allow the UAM Coordinator to monitor timestamped actions (initiated by DUC or UAM Coordinator), and U-space configurations, abnormal situations, and conformance issues, fostering shared awareness, oversight, and transparency between the UAM Coordinator and DUC.

A primary high-level task is for the UAM Coordinator and DUC to effectively manage emergency situations. The purpose of the emergency management task is to:

1. assist U-space users (i.e., UAS/UAM operators) experiencing an emergency to manage the emergency (e.g., report the emergency, find solutions, activate contingency plans);
2. provide information to others (who may be concerned/involved in assisting the emergency); and
3. adhere to compliance with emergency procedures (e.g., configure dynamic safety boundaries, change flight prioritisation, plan and coordinate emergency routings).

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

This process necessitates close coordination between the UAM Coordinator, UAS/UAM operators, and emergency services (e.g., SOS, police). Checklists will be used to support the UAM Coordinator and DUC in responding to emergency situations, with DUC presenting and marking completed tasks. The checklist includes items like retrieving U-plans (with contingency plans), assessing risks, and identifying nearby vehicles. Both the UAM Coordinator and DUC maintain continuous monitoring of emergency communication channels, such as aviation's 121.5 frequency.

Performance Targets

The following table presents metrics that can be used during evaluations, at different stages of development. Initially, we aim for a high-level assessment of the viability of the concept, with these KPI:s in mind.

Table 39. *UC3 Performance Targets*

KPA	Category	KPI
System Performance	MoP	Task completion time, task accuracy, system availability, timeliness, synchronisation, and data exchange accuracy.
Human Performance	MoP	Usability, workload, understanding, situational awareness, operational method, acceptance, and trust.
Safety	MoE	Adherence to safety procedures, lateral/vertical separation, and airspace restrictions.

High Level Validation Requirements

For VAL2 in UC3, the following high level validation applies.

Table 40. *HAIKU_UC3_HLR_1*

HL-REQ-ID	HAIKU_UC3_HLR_1:
Requirement	The DUC must provide relevant and real-time information to support the UAM Coordinator's high-level decision-making, including traffic monitoring, flight information and emergency response plans.
Rationale	Providing relevant and real-time information to the UAM Coordinator supports their decision-making process. The DUC consolidates data from multiple sources, such as traffic monitoring and flight information, enabling the UAM Coordinator to make informed decisions based on accurate and up-to-date information.
KPA	System Performance, Human Performance, 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

Table 41. *HAIKU_UC3_HLR_2*

HL-REQ-ID	HAIKU_UC3_HLR_2:
Requirement	The DUC shall effectively automate standard and repetitive tasks to free up time and effort required by the UAM Coordinator to focus on high-level tasks.
Rationale	Automating standard, repetitive tasks through the DUC reduces the burden on the UAM Coordinator, allowing them to focus on higher-level decision making. By streamlining routine operations, the DUC enhances operational efficiency, improves productivity, and reduces the potential for human error.
KPA	System Performance, Human Performance, Safety

Table 42. *HAIKU_UC3_HLR_3*

HL-REQ-ID	HAIKU_UC3_HLR_3
Requirement	The team must be able to manage emergency situations safely and expeditiously.
Rationale	Safe and expeditious management of emergencies requires a clear division of roles and responsibilities for UAM Coordinator and DUC. The objective is to structure teamwork in emergency situations using a checklist approach where the DUC and UAM Coordinator collaborate according to a checklist.
KPA	System Performance, Human Performance, Safety

Table 43. *HAIKU_UC3_HLR_4*

HL-REQ-ID	HAIKU_UC3_HLR_4
Requirement	The team must be able to solve U-plan issues (on a pre-tactical or tactical timescale) by developing and activating new routes and conducting required coordination with affected stakeholders.
Rationale	To ensure that traffic is safely separated, the team must be able to detect U-plan issues (e.g., no-fly zones) and rerouted traffic according to U-space constraints.
KPA	System Performance, Human Performance, Safety

Table 44. *HAIKU_UC3_HLR_5*

HL-REQ-ID	HAIKU_UC3_HLR_5
Requirement	The team must be able to solve U-plan conflicts between aircraft (on a pre-tactical or tactical timescale).

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

Rationale	To ensure that traffic is safely separated, the team must be able to detect traffic conflicts and resolve these as needed.
KPA	System Performance, Human Performance, Safety

Table 45. *HAIKU_UC3_HLR_6*

HL-REQ-ID	HAIKU_UC3_HLR_6
Requirement	The team must be able to monitor the movements of all aircraft in the U-Space and determine non-compliance with their U-plans.
Rationale	To ensure U-space safely, the team must be able to track flights and detect U-plan compliance issues (i.e., deviations) and take actions to resolve the situation.
KPA	System Performance, Human Performance, Safety

Table 46. *HAIKU_UC3_HLR_7*

HL-REQ-ID	HAIKU_UC3_HLR_7
Requirement	The team must be able to receive and process information on ground activities that may affect the U-space.
Rationale	To ensure U-space safely, the team must be able to configure the U-space dynamically according to ground events (e.g., fires generating smoke that constrain a portion of the U-space).
KPA	System Performance, Human Performance, Safety

Table 47. *HAIKU_UC3_HLR_8*

HL-REQ-ID	HAIKU_UC3_HLR_8
Requirement	The team must be able to handle an amount of airborne traffic representative to future UAM scenarios. Note that this HLR is defined in relation to timelines in U-space and UAM ConOps descriptions as opposed to absolute traffic numbers.
Rationale	Higher traffic numbers, and a more complex U-space, drive the need for intelligent assistant capabilities and humans alone will not be able to handle the U-space independently.
KPA	System Performance, Human Performance, 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

4.4 UC#3 Second Validation Plan (VAL2)

VAL2 will explore all the objectives described in Table 38. Building on the validation plan in VAL1, VAL2 will explore simulation-based testing.

Validation Objectives

The HAT objectives in UC3 have been defined to address four key human factors challenges for Human-AI Teaming: situation awareness, transparency, decision-making, and bi-directional communication. In relation to these challenges have been articulated as capabilities that DUC should possess to support collaboration.

Situation awareness: The situation awareness objective requires that DUC continuously monitor U-space and traffic operations, processing real-time data from various sources to detect trends (e.g., traffic congestions, flight patterns due weather), anomalies, potential conflicts (that DUC the resolves), and anticipate future traffic patterns. Additionally, DUC should effectively communicate with the UAM Coordinator by providing real-time updates and status reports on the traffic situation and U-space services, responding to information requests, inferring information needs, and guiding the UAM Coordinator's attention to critical events.

Transparency: The transparency objectives require that DUC provides clear and relevant explanations upon the UAM Coordinator's request, accurately discerning the Coordinator's informational needs. Furthermore, DUC must demonstrate the significance of its explanations for specific decisions or actions while adapting the level of abstraction based on the task, situation, and the UAM Coordinator's expertise. Additionally, DUC should articulate how it generated its outputs and explain its operational processes.

Bi-directional communication: The bi-directional communication objectives require DUC to acknowledge the UAM Coordinator's instructions and intentions effectively. It must possess the capability to understand and generate natural language while utilising various communication modalities, such as voice, text, and graphics, to enhance clarity. Additionally, DUC should refrain from interrupting the UAM Coordinator with voice or aural messages when the UAM Coordinator is busy with other communications or activities. DUC must be able to automatically adapt its interaction modalities based on the user's preferences, states, and situational context.

Decision making: The decision-making objectives require that DUC be capable of solving conflicts, recommending actions and solutions to the UAM Coordinator while making and executing decisions within its operational parameters. Furthermore, DUC must identify suboptimal strategies proposed by the UAM Coordinator and suggest alternatives with appropriate justifications. Additionally, it should be capable 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

problem-solving in collaboration with the UAM Coordinator using a checklist approach (e.g., is response to emergency situations (e.g., a sick passenger in an Air Taxi requiring medical assistance: Event 1).

The HAT objectives outlined above have been aligned with the EASA HAT requirements as documented in D4.1 (HAIKU, 2024). The tables below indicate the relevant EASA HAT requirements from D4.1 (HAIKU, 2024) that will be examined in VAL2, detailing the corresponding validation objective, activity, and metric.

EASA HF Requirements

Table 48. *UC3 Cooperation/collaboration capabilities (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
HF-01: DUC must be able to provide information about which elements in the environment that are monitored and distinguish between normal and emergency situations.	NO			
HF-02: DUC must be able to guide the UAM Coordinator's attention to important information based on an understanding of where the UAM Coordinator is at the moment (see HF-13)..	YES	To explore the extent to which DUC can direct the human's attention to an event. To explore the extent to which DUC can determine where the human's attention is. To evaluate subjective situation awareness. To evaluate subjective workload. To evaluate functionality of attention guidance. To explore the advantages and disadvantages of using eye-tracking to determine visual focus and guide attention.	Eye-tracker data logs & questionnaire	Quantitative & qualitative. E.g., How long does it take to orientate to the information/ event on the map.

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

HF-02: DUC must be able to provide information about specific U-space elements and flights on the UAM Coordinator's request.	YES	To measure problem solving time. To evaluate subjective situation awareness. To evaluate subjective workload. To evaluate and determine an appropriate level of information detail to be provided by DUC regarding U-space elements (geo-fence and no-fly zones, vertiports and cargo hubs, U-space constraints) and aircraft positions and intentions (e.g., flight plan).	Simulator data logs, questionnaire and de-briefing.	Qualitative. E.g., How complete is the situation? To what extent is the information valuable to solve the situation?
HF-03: DUC must be able to track the UAM Coordinator's visual attention to identify gaps in his/her perception of elements in the environment.	YES	To evaluate the ability to track end users' visual attention on the situation display. To evaluate subjective situation awareness. To evaluate subjective workload. To evaluate functionality of attention tracking.	Eye-tracker data logs, questionnaire, de-briefing..	Quantitative and qualitative.
HF-04: DUC must allow the user (UAM Coordinator) the flexibility to accept, reject, amend, and/or ask for an explanation, in response to any provided recommendation.	YES	To evaluate the usability of the IA's dialogue windows. To evaluate subjective satisfaction of Human-IA interaction (with dialogue windows). To evaluate functionality of dialogue windows with recommendations. To evaluate subjective decision making processes.	Observations, questionnaire, de-briefing.	Qualitative

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

HF-04: The response time window available (to accept / reject / amend / seek explanation) must be adequate to allow the user to assess the recommendation.	YES	To evaluate the subjective efficiency in responding to recommendations made by the IA. To evaluate functionality of dialogue windows with recommendations. To evaluate subjective satisfaction of Human-IA interaction (with dialogue windows).	Simulator data logs, observations, questionnaire, de-briefing.	Quantitative and qualitative.
HF-04: System response logic (time-out vs auto-implement) shall be made explicit, in response to user actions (i.e. accept / reject / amend / seek explanation) .	YES	To evaluate subjective situation awareness. To evaluate subjective satisfaction of Human-IA interaction (with dialogue windows). To evaluate the usability of the IA's dialogue windows.	Questionnaire, de-briefing.	Qualitative.
HF-08: DUC must be able to suggest alternative solutions and provide arguments supporting them.	YES	To evaluate functionality of Human-IA interaction (with dialogue windows). To evaluate subjective satisfaction of Human-IA interaction (with dialogue windows). To evaluate the usability of the IA's dialogue windows.	Simulator data logs, questionnaire, de-briefing.	Qualitative.
HF-09: The UAM Coordinator must be able to adjust high-level task parameters for DUC, such as adjusting goals and constraints for capacity thresholds, conformance monitoring, and alerts and warnings.	YES	To evaluate functionality of Human-IA interaction (IA high-level task parameters). To evaluate subjective satisfaction of Human-IA interaction (IA high-level task parameters). To evaluate the usability of adjusting IA high-level task parameters.	Simulator data logs, questionnaire, de-briefing.	Qualitative.

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 49. *UC3 Speech / Natural Language (EASA Macroarea)*

EASA-like requirement (D4.1)	**Covered**	**Validation Objective**	**Activity**	**Metric**
HF-10: DUC must be able to propose the correct (normal- or emergency), or provide it as requested by the UAM Coordinator.	YES	To evaluate subjective satisfaction of Human-IA interaction (communication). Spoken language communication achieved through Wizard of Oz.	Questionnaire, de-briefing.	Qualitative.
HF-11: DUC must alert the user to any potential user misunderstandings, as inferred by the UAM Coordinator's visual attention (see HF-12)	YES	To evaluate subjective satisfaction of Human-IA interaction (communication). Spoken language communication achieved through Wizard of Oz. To evaluate functionality of Human-IA interaction (communication).	Eye-tracker data logs, Questionnaire, de-briefing.	Qualitative.
HF-12: DUC must be able to determine if the UAM Coordinator has misunderstood an alert provided by DUC based on the UAM Coordinator's visual attention.	YES	To evaluate subjective satisfaction of Human-IA interaction (communication). Spoken language communication achieved through Wizard of Oz. To evaluate functionality of Human-IA interaction (communication).	Eye-tracker data logs, Questionnaire, de-briefing.	Qualitative.
HF-13: If DUC provides information or an alert related to a situation that is shown on the map display, DUC can assess if the UAM Coordinator has understood the information or alert by checking if the UAM Coordinator is looking at the correct location on the HMI. If the UAM Coordinator is not looking at the correct location, DUC should be able to guide the	YES	To explore the extent to which DUC can direct the humans attention to an event. To explore the extent to which DUC can determine where the human's attention is. To evaluate subjective situation awareness. To evaluate subjective	Eye-tracker data logs & questionnaire	Quantitativ e & qualitative. E.g., How long does it take to orientate to the information /event on the map.

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

UAM Coordinators attention (see HF-02).		workload. To evaluate functionality of attention guidance. To explore the advantages and disadvantages of using eye-tracking to determine visual focus and guide attention. Note: Spoken language communication achieved through Wizard of Oz manipulation.		
HF-14: DUC must not "step on" other ongoing user communications	YES	To evaluate subjective satisfaction of Human-IA interaction (communication). Spoken language communication achieved through Wizard of Oz manipulation.	Questionnaire and de-briefing.	Qualitative.

Table 50. *UC3 Multimodal Interaction (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
HF-21: DUC must permit user deactivation of voice mode	YES	To evaluate subjective satisfaction of Human-IA interaction (communication).	Note interaction choice, questionnaire and de-briefing.	Qualitative.
HF-24: : DUC must allow adaptation of its notification modalities (visual, aural) depending on situation state (nominal / non-nominal / emergency) and user cognitive state.	YES	To evaluate subjective satisfaction of Human-IA interaction (communication). To evaluate subjective situation awareness.	Questionnaire and de-briefing.	Qualitative.

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 explore the advantages and disadvantages of different notification modalities.		
HF-25: DUC must allow adaptation of its interaction modalities depending on situation state (nominal / non-nominal / emergency)	YES	To evaluate subjective satisfaction of Human-IA interaction (communication). To evaluate subjective situation awareness. To explore the advantages and disadvantages of different interaction modalities.	Questionnaire and de-briefing.	Qualitative .

Table 51. *UC3 Error Management (EASA Macroarea)*

EASA-like requirement (D4.1)	**Covered**	**Validation Objective**	**Activity**	**Metric**
HF-28: DUC must have a clear and intuitive HMI that minimises confusion and misunderstanding of how to interact with DUC.	YES	To evaluate subjective satisfaction of Human-IA interaction (communication). To analyse error patterns (identify common mistakes or errors made by participants).	Observations, questionnaire and de-briefing.	Qualitative.

XAI Requirements

Table 52. *UC3 Attributes of OpXAI (EASA Macroarea)*

EASA-like requirement (D4.1)	**Covered**	**Validation Objective**	**Activity**	**Metric**
EXP--10: DUC must be able to explain its decisions and recommend stations using progressive CLT levels, for its recommended actions. regarding, traffic monitoring, coordination with vertiports, and provision of flight and weather information.	YES	To evaluate subjective satisfaction of explanations. To investigate advantages and disadvantages of applying CLT theory for presenting explanations To explore the advantages and disadvantages of using storytelling as a method for presenting explanations.	Observations, questionnaire and de-briefing.	Qualitative

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

EXP--11: DUC explanations must be presented using a combination of text, symbols, and graphical overlays on the map display	YES	To evaluate subjective satisfaction of explanations. To explore the advantages and disadvantages of different presentation techniques. To evaluate functionality of storytelling explainer, linked to situation display, for presenting explanations.	Observations, questionnaire and de-briefing.	Qualitative
EXP--12: DUC should be able to explain why different solution alternatives are proposed, including factors considered and how relevant they are, and the underlying decision making process.	YES	To evaluate subjective satisfaction of explanations. To explore the advantages and disadvantages of using storytelling as a method for presenting explanations.	Observations, questionnaire and de-briefing.	Qualitative
EXP--13: DUC must present its decision making rationale both prior to, and during, situations, consistent with CLT. Prior explanations shall rely on a storytelling approach, concurrent explanations shall use a combination of text, symbols, and graphical overlays on the map display	YES	To evaluate subjective satisfaction of explanations. To investigate advantages and disadvantages of applying CLT theory for presenting explanations To explore the advantages and disadvantages of using storytelling as a method for presenting explanations.	Observations, questionnaire and de-briefing.	Qualitative.
EXP--14: DUC must provide dynamic adaptation of explanation levels, per CLT	YES	To evaluate subjective satisfaction of explanations. To investigate advantages and disadvantages of applying CLT theory for presenting explanations To explore the advantages and disadvantages of using storytelling as a method for presenting explanations.	Observations, questionnaire and de-briefing.	Qualitative.
EXP--15: Explanations prior to a situation (for training purposes) are timed to appear during periods with little	YES	To evaluate subjective satisfaction of explanations.	Observations, questionnaire	Qualitative.

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

activity. Explanations for decisions/actions in time-critical situations are timed to occur when DUC has derived a solution. The explanation is valid as long as the proposed decision/action is valid.		To investigate advantages and disadvantages related to the timing for presenting explanations.	and de-briefing.	
EXP--16: The system must retain a user-selectable explanation level (according to CLT), for all explanations provided by DUC.	YES	To evaluate subjective satisfaction of requesting explanations. To evaluate functionality of requesting explanations according to different CLT levels.	Observations, questionnaire and de-briefing.	Qualitative.
EXP--17: The system must provide the user valid and reliable explanations, consistent with information obtained from all users, as well as inputs from weather, traffic, emergency services, and ground services sources.	NO			

Validation Approach

We plan to conduct simulator trials with traffic management and coordination domain experts (i.e., air traffic controllers, flow managers, Joint Rescue Coordination Center Rescue Leaders) acting as the UAM Coordinator to simulate traffic scenarios in a fictional U-space for Stockholm, Sweden. Scenarios will include several events that participants are asked to resolve together with DUC. In VAL2, we will focus more on implementing and validating DUC capabilities and the DUC HMI. Additional scenarios will be developed and simulated to study how participants, acting as the UAM Coordinator, perceive and interact with DUC to accomplish various high-level tasks, with a particular emphasis on emergency management. By concentrating on emergency management, the DUC concept will demonstrate its ability to address various operational aspects, ensure safety, integrate with existing systems, and enhance user experience. Furthermore, we will focus on implementing DUC capabilities related to situation awareness, decision-making, transparency, and bi-directional communication, either in a prototype or through a Wizard of Oz approach.

Participants will first receive an introduction to the research and simulation study in which they are participating, during which informed consent will be collected.

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

Following this, they will receive a briefing on the simulator, their working position, their roles, and their tasks. This will be followed by a training session in which they will engage in a scenario and interact with DUC. Afterward, they will participate in the simulation exercise, during which and afterward dependent measures will be collected. A post-simulation debriefing will conclude the study.

Exercise Description and Scope

The HAT for VAL2 will be tested.

Table 53. *UC3 Exercise description*

Exercise Description	Scope (What is being assessed)	Key R&D Objectives explored
Emergency management. Rerouting of flight due to medical emergency	Collaboration, decision-making, usability, acceptability, trust, situation awareness, explainability	UC3-OBJ-01/02/03/04/05
Emergency management Dynamic capacity management and rerouting due to Vertiport on fire.	Communication effectiveness, collaboration, decision-making, usability, acceptability, trust, situation awareness, explainability	UC3-OBJ-01/02/03/04/05
Emergency management. Conformance issue and attention guidance caused by simultaneous events occurring, with flight deviating from U-plan and experiencing link loss.	Attention guidance, collaboration, decision-making, usability, acceptability, trust, situation awareness, explainability	UC3-OBJ-01/02/03/04/05

Exercise Scenarios

One or more tailored simulation scenarios will be developed that capture the essential aspects of DUC and UAM Coordinator teaming facing emergency management situations. These scenarios will include one or more events for participants to solve with the assistance of DUC. The following scenarios in the table below are candidates for VAL2.

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 54. *UC3 Exercise Description and scenarios*

Scenario description	Short description	Key R&D Objectives explored
Event 1: Single passenger air taxi transporting an unresponsive passenger from Globen to Karolinska University Hospital in Stockholm Sweden (used in VAL1)	High-level task: Emergency management. A passenger travelling in an uncrewed air taxi becomes sick. The UAM Coordinator is informed of the sick passenger. The passenger requires immediate treatment and a decision must be made on how to best accomplish that.	UC3-OBJ-01/02/03/04/05
Event 2: Vertiport on fire requiring closure of vertiport and a portion of the U-space airspace.	High-level task: Emergency management and dynamic capacity management. The closure of the U-space volume and vertiport triggers a need for dynamic capacity management where affected traffic is rerouted or delayed.	UC3-OBJ-01/02/03/04/05
Event 3: Conformance monitoring issue with flight deviating from U-plan with subsequent link loss (lost tracking).	High-level task: Conformance Monitoring and Emergency management. DUC highlights the UAM Coordinators attention to a conformance monitoring issue with a flight deviating from its flight plan. The flight then disappears from the situation display, which triggers an emergency procedure and initiation of search and rescue.	UC3-OBJ-01/02/03/04/05

Participants and their Role

For VAL2, air traffic controllers and flow managers will participate in simulations planned to be held at Malmö ATCC, Sweden. The target is 10 participants. In addition, we seek to invite operators from other traffic management domains, such as train traffic control, and coordination domains, such as Rescue Leaders from the Joint Rescue Coordination Center (JRCC).

Participants would act as the UAM Coordinator in managing the U-space over Stockholm City. Prior to running the simulation, participants will receive training on the role, simulator, and interaction with DUC.

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

Platform / Tools & Technique

The following simulator and tools are planned to be used in VAL2.

A setup of the UAM Coordinator working position consisting of:

- Screen 1: Knowledge library with manuals, checklists, procedures to guide work in different situations. DUC can suggest a checklist to apply for different situations. For this we plan to use a portable computer connected to an external screen.
- Screen 2: Situation display with a map view of the U-space using UTM City. The situation display will show U-space users (i.e., traffic) and U-space elements (e.g., vertiports, geo-zones, no-fly zones etc). For this we plan to use a portable computer connected to an external screen.
- Screen 3: Storytelling explainer and Digital journal/logbook system, and communication log (e.g., DUC, UAM Coordinator, UAS/UAM operators, vertiport operators, other stakeholders). Voice communication is transcribed and logged. For this we plan to use a portable computer connected to an external screen.
- Communication system: Phone to contact UAS/UAM operators and vertiport operators. UAM Coordinator can request the DUC to initiate calls.

Traffic flows and scenario events will be scripted in the scenario builder.

The DUC prototype is planned to be implemented using a combination of DUC HMI overlays on the situation display and Wizard of Oz (e.g., triggering events, communications, interactions).

We plan to make use of eye tracking to monitor participants' visual attention and provide attention guidance to areas on the situation display that require the UAM Coordinator's attention.

Intelligent Assistant

DUC aids the UAM Coordinator in overseeing U-space operations. It handles most routine tasks, such as, traffic tracking, conformance monitoring, and providing flight and weather information. Additionally, the DUC directs the UAM Coordinator's attention to specific situations or events as necessary, using visual cues in the interface.

Furthermore, the DUC supports the UAM Coordinator by:

- Assisting UAM vehicle operators in emergencies by providing reports, suggesting actions, and offering contingency plans.
- Following emergency procedures, such as adjusting dynamic safety boundaries, changing flight prioritisation, and coordinating emergency routes.

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

- Establishing priority criteria dynamically for different types of flights.
- Providing the human with information about the situation – situation awareness
- Facilitating the space needed for the human to coordinate with other stakeholders to find a solution – reduce workload.

Data Collection and Analysis Methods

During the simulations, participants will interact with DUC to solve various tasks and problems related to different normal events and emergency events. The simulations will assess the teamwork between the DUC and the UAM Coordinator, focusing on participants' perception of DUC and how well it supports situation awareness, decision-making, understanding, and bi-directional communication.

To gather valuable insights, feedback will be collected from participants acting as the UAM Coordinator involved in the simulations. This feedback will cover aspects such as technical feasibility and functionality of the system, and the end users' understanding, and experience of working with the AI as a Digital Assistant. For this purpose, a combination of observations, questionnaires, interviews, and system logs will be collected and analysed.

Planned Activities

Table 55. *UC3 Planned Activities*

Activity	Activity description	General information
VAL2 Strategy and Plan	Define VAL2 Strategy and Plan	Deadline 31st Oct.. 2024
Scenario Design and Implementation	Script scenarios in detail and then implement them in the simulator.	Deadline 31st Oct. 2024
ConOps iteration.	Iterate and implement a simulator environment with ConOps updates to Stockholm U-Space, UAM Coordinator, and DUC. This includes the UAM Coordinator working position and emergency management checklists.	Deadline 31st Dec. 2024

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

HMI Development	Iterate HMI (UTM City, Storytelling Explainer, knowledge library and checklists) to meet Scenario and ConOps requirements.	Deadline 31st Dec. 2024
System integration	Finalise simulator for VAL2. Connect all components into a working simulator.	Deadline 31st Jan 2025
Update VAL2 Strategy and Plan	Update strategy and plan to harmonise with capabilities of simulator and scenarios.	Deadline 31st Jan 2025
Preparation and 3 Tests Runs	Testing of simulator and validation procedures prior to validation.	Deadline 28th Feb. 2025.
VAL2	Evaluate DUC concept and DUC and UAM Coordinator teamwork.	Deadline 31st Mar. 2025

Table 56. *UC3 GANTT*

	ConOps & Requirement Definition
	Design & Development
	Validation

Year	**2024**						**2025**							
Month	7	8	9	10	11	12	1	2	3	4	5	6	7	8
Human-AI Teaming concept														
AI-based system component/s														
Non-AI technological component/s														
HMI														

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 Use Case #4 – Digital and Remote Tower

The Intelligent Sequence Assistant (ISA) aims to support and enhance decision-making for Air Traffic Controllers by optimising runway utilisation in single-runway airports. It provides sequence suggestions for both arriving and departing aircraft, aiming to streamline operations and improve efficiency.

5.1 UC#4 TRL Overview

Table 57. *UC4 TRL Overview*

COMPONENTS: Optimisation Algorithm to output sequence, Service for computing ETA and initial trajectory points		
TRL	**Month**	**Activity to reach selected level**
1	1	Concept formulated at the project's start.
2	12	Literature review and meetings with operational experts to refine the concept.
3	18	An initial proof of concept of the technology was discussed. There were both technical feasibility and desired features assessments. Before the VAL1 experiment, data was shared among the technical partners and an initial prototype was developed to test the models and the new features
4	24	VAL1: Sequencing exercises were conducted by ATCOs, and their results were subsequently compared to those generated by the ISA for the same exercises. The results were found to be comparable, leading to the transition to TRL 4
5	30	The aim is to progress from TRL 5 from VAL1 to VAL2 by validating each component individually in a relevant environment, specifically using the Alicante airport scenario in the simulator.
6	33	The aim is to reach TRL 6 after VAL2, as the system will be validated as a whole

COMPONENTS: Explainability for sequence changes, HMI - Electronic flight strips and strip board management		
TRL	**Month**	**Activity to reach selected level**
1	1	Concept formulated at the project's start.

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	12	Literature review, meetings with Experts, user needs specifications gathered through user research (i.e.: Observational studies in the Tower by Human Factors experts).
3	24	VAL1: we designed a realistic, interactive mock-up of the HMI with the electronic strips and the explanations. These were tested qualitatively with users, who validated our design.
4	28	Full prototype of each component developed and tested.
5	30	The aim is to progress from TRL 5 from VAL1 to VAL2 by validating each component individually in a relevant environment, specifically using the Alicante airport scenario in the simulator
6	33	The aim is to reach TRL 6 after VAL2, as the system will be validated as a whole

Another key component of the architecture of ISA is the Core Processing Module. This module acts as a connector between all other components and manages their interactions. Its TRL level corresponds to that of the entire system, and therefore an analysis of its individual evolution is unnecessary.

5.2 UC#4 Validation 1 Results and Lessons Learned

The completion of the VAL1 experiment was one of the main steps regarding the TRL evolution. These were the main insights which led to a ConOps Update.

Table 58. *UC4 VAL1 lesson learned and new design directions*

Insight	**ConOps Update**
Asynchronous sequencing works to a satisfactory level and aligns with ATCO's mental models. However, the validation suggested that ATCOs would need a real-time sequence. Without it, ISA would not be operationally viable.	ISA must be able to operate in real-time to adapt to real operations
ISA's scope is too large at the moment, as the sequence suggested may be too long	ISA should focus only on aircraft which are going to use the runway very soon

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

ISA is not useful to ATCOs in easy situations (when aircraft are lined up easily, one after the other). It is much more useful in tense situations when sequencing calculations are tricky (such as when you have three arrivals and two departures one after the other).	ISA should help ATCOs solve challenging sequences and should not constantly attract ATCOs' attention when easy sequences are involved.
ISA's control settings are a bit too generic	ISA's settings must take into account Alicante's airport specific regulations, and also scenarios where ISA must be turned off to not create problems
CLT1 (Overview of operation) is confusing: ATCO could not tell whether it was prospective or retrospective information	CLT1 should provide a mental picture of what's about the happen, and should not create confusion
CLT3 (Sequence change) was not understood fully	CLT3 should provide an on-demand detailed explanation for the sequence change that matches ATCO's way of thinking
CLT5-6 were considered useless	CLT5-6 should provide the most detailed explanations for a debriefing post-operations, but it should not be the same as data provided by EUROCONTROL

5.3 UC#4 Updated Concept of Operations

The insights gathered from VAL 1 directly led to improvements in the ConOps and to the design of the IA. Below are the main changes that will be present in VAL 2:

- ISA will now operate during real-time operations, pending successful integration with current simulator's systems.
- ISA will only display three aircraft in the sequence (the next 3 to use the runway).
- The interface of CLT has been improved in terms of usability and organisation of information displayed.
- CLT5-6 were ultimately dropped, as the huge amount of information is not useful during operations.

Key R&D Needs and Objectives

The key objectives to explore and validate 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

1. ATCO must be able to understand the logic of the sequence suggested by ISA.
2. ISA must be able to recalculate the sequence every time there is an input (new action, new aircraft, emergency, etc) every second.
3. The team (ATCO and ISA) must decrease the ATCO's subjective workload during periods of high demand.
4. The team (ATCO and ISA) must always keep situational awareness.

Problem Statement - what is the problem to address with the Intelligent Assistant?

Intelligent Assistants would imply a significant increase of capacity and provision of new safety nets. Such assistants could be designed to aid air traffic controllers in a wide variety of tasks. For instance, provision of air traffic control in single runway airports is challenging, since the same runway is used for both landings and take-offs. The capacity of such airports is normally limited by runway occupancy time, and thus to maximise capacity one needs to optimise runway throughput by optimising aircraft sequencing. A Intelligent Assistant that aided the controller in tasks such as vacating the runway and overflying the runway end would significantly reduce workload and ensure that the best possible sequence is designed by the tower controller, so as to maximise runway throughput.

Human Factors & Safety Risks – Potential risks and problems introduced by the Intelligent Assistant

- **Situation awareness:** the situation awareness of ATCOs would be disrupted if ISA suggested sequences that they did not think about.
- **Human performance:** human performance would be impacted negatively if ATCOs do not understand properly what ISA is doing, especially at the beginning as they start to get acquainted with the system
- **Trust:** Since all ATCOs have their own experience and preferred sequencing style, they may not fully trust the suggestions provided by ISA. This could lead to ATCOs disregarding ISA's safest solutions or even disabling it entirely, even during critical moments like peak traffic periods when its assistance would be most valuable.

R&D Objectives

Table 59. *UC4 Key R&D Objective*

OBJ-ID	Validation Objective	Success Criteria ID	Success Criteria
UC4-OBJ-01	To assess the Operational Feasibility Acceptability of the solution from the ATC	UC4-CTR-01	The solution is considered operationally feasible and

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

	perspective in nominal conditions.		acceptable by TWR ATCOs in nominal conditions.
UC4-OBJ-02	To assess the solution Safety-wise from the ATC and Safety Team perspective in nominal conditions.	UC4-CTR-02	The solution is considered acceptable by TWR ATCOs and the Safety Team if it doesn't introduce additional safety risks to operations.
UC4-OBJ-03	To demonstrate that the solution has a high trust factor	UC4-CTR-03	The solution is considered trustworthy in nominal conditions.
UC4-OBJ-04	To assess the ISA HMI interface and information requirements related to explainability.	UC4-CTR-04	ATCOs give positive feedback on the current status of the HMI and the explainability provided.

Operational Concept Description

The ISA is an AI-based system designed to optimise the sequencing of arriving and departing aircraft at single-runway airports. The system integrates real-time data, computes optimal sequences, and provides actionable information to air traffic controllers through an intuitive HMI. The primary goals are to enhance runway utilisation, ensure safety, and improve operational efficiency.

Task Allocation Pattern

The task allocation between the end users (Air Traffic Controllers) and the AI-based system (ISA) is designed to leverage the strengths of both human expertise and advanced AI capabilities. Here's a detailed task allocation pattern:

A. Air Traffic Controllers:

- Goals and High-level tasks:
    - Decision-making for runway usage based on AI-generated sequences.
    - Monitoring and adjusting sequences in response to real-time conditions and system alerts.
    - Handling exceptions and emergencies with priority over automated suggestions.
- AI System Interaction:
    - Receive sequence suggestions from ISA.

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

  - Provide real-time feedback to ISA for continuous learning and adjustment.
  - Utilise HMI to update aircraft status and manage the strip board.

B. AI-Based System (ISA)

- Primary Responsibilities:
  - Compute optimal sequences for arrivals and departures using real-time data and machine learning algorithms.
  - Continuously integrate traffic surveillance, meteorological data, and runway status to update sequences.
  - Provide real-time alerts and recommendations through the HMI.
- Human Interaction:
  - Receive feedback from controllers to refine and adjust the sequencing algorithms.
  - Provide intuitive and actionable information to controllers for decision support.

Operational Definition (OD)

- Scope:
  - The system is designed for single-runway airports to enhance efficiency and safety.
  - Focuses on both arrival and departure sequencing, taking into account real-time data and operational constraints.
  - Designed to always suggest the next 3 aircraft to use the runway (maximum).
- Operational Environment:
  - Developed in a simulator that replicates Alicante airport.
- Performance Metrics:
  - Runway utilisation rate.
  - Average delay times for arrivals and departures.
  - Safety incidents related to sequencing and spacing.

Specific Operational Limitations and Assumptions

- Limitations:
  - The system relies heavily on accurate and timely data from traffic surveillance and meteorological sources.

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

    - Unexpected events such as sudden weather changes or emergencies require human intervention and may limit the system's effectiveness.
    - The effectiveness of ISA is contingent on the controllers' acceptance and correct interpretation of the AI-generated suggestions.
- Assumptions:
    - Controllers are fully trained and proficient in using the ISA system and HMI.
    - Real-time data feeds are reliable and continuously available.
    - The airport operates under standard operational procedures, with predefined protocols for handling exceptions and emergencies.

Detailed ConOps Description

1. Pre-Operation Setup:

- Briefing: Controllers receive a pre-ops briefing.
- System Initialization: ISA initialises, pulling the latest data from traffic surveillance and meteorological systems.

2. Operational Phase:

- Data Integration: Real-Time Data Integration Subsystem (RTDIS) continuously feeds real-time data into the system.
- Sequence Computation: Sequence Calculation Module (SCS) calculates optimal sequences, updating them as new data comes in.
- HMI Interaction: Controllers interact with the HMI, reviewing and adjusting sequences as necessary.
- Feedback Loop: Controllers provide feedback on sequence suggestions, which ISA use to refine its algorithms.

3. Exception Handling:

- Emergency situations and unexpected events: Controllers take immediate manual control, disabling ISA recommendations (no sequence needed as traffic in distress will always be number 1).

4. Post-Operation Analysis:

- Performance Review: Post-operation, ISA generates performance reports,

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

highlighting areas of success and potential improvement.
- Controller Feedback: Controllers provide detailed feedback on system performance, which is used for future updates and training.

The ConOps for the AI-based ISA system ensures a balanced task allocation between advanced AI capabilities and human expertise. By clearly defining the roles, operational scope, and limitations, the system aims to enhance air traffic management at single-runway airports, ensuring safety, efficiency, and optimal runway utilisation.

Performance Targets

Table 60. *Key Performance Areas*

KPA	Category	KPI
Operational, Traffic Management Efficiency	MoE	Number of arrivals and departures managed in an hour/X minutes.
Safety	MoE	Number of safety events in an hour/X minutes (including "go around" manoeuvres, aborted take-off, etc.).
Human Performance	MoP	Conformance (Comparison between "ideal" sequence vs. the sequence carried out by the ATCO); Trust factor; Usability; User Experience; Workload (subjective assessment).

High Level Validation Requirements

Table 61. *HAIKU_UC4_HLR_1*

HL-REQ-ID	HAIKU_UC4_HLR_1:
Requirement	The team must keep situational awareness of the traffic situation at all times.
Rationale	The team must always keep situational awareness of the traffic situation.
KPA	Operational, Safety, Human Performance

Table 62. *HAIKU_UC4_HLR_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

HL-REQ-ID	HAIKU_UC4_HLR_2:
Requirement	The team must maximise company policy operational goals.
Rationale	The team must maximise company policy operational goals.
KPA	Operational, Safety, Human Performance

Table 63. *HAIKU_UC4_HLR_3*

HL-REQ-ID	HAIKU_UC4_HLR_3:
Requirement	The team decision must represent the best trade-off between company goals and ATCO's feedback.
Rationale	The team decision must be based on efficiency (e.g. maximise the throughput of arrivals and departures in an hour) and ATCO's feedback. Most efficient solution might not be feasible since it could mean assuming too much risk when managing a sequence of consecutive arrivals and departures. The ideal solution would be a system that can be tuned to offer a more aggressive or more conservative sequence, depending on the situation/ATCO in charge, and that could maximise the number of arrivals over departures (or the other way around).
KPA	Operational, Safety, Human Performance

Table 64. *HAIKU_UC4_HLR_4*

HL-REQ-ID	HAIKU_UC4_HLR_4:
Requirement	The team must maintain adequate situational awareness.
Rationale	The team must be able to keep situational awareness of safety events.
KPA	Safety

Table 65. *HAIKU_UC4_HLR_5*

HL-REQ-ID	HAIKU_UC4_HLR_5:
Requirement	The team must achieve a "high trust factor."
Rationale	"High trust factor" means most ATCOs trust IA suggested sequence.
KPA	Human Performance

Table 66. *HAIKU_UC4_HLR_6*

HL-REQ-ID	HAIKU_UC4_HLR_6:

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

Requirement	The team must increase runway throughput by an average of 5% (number of aircraft per hour).
Rationale	Higher throughput means larger volume of operations that the Alicante airport can handle and, therefore, higher efficiency
KPA	Operational

Table 67. *HAIKU_UC4_HLR_7*

HL-REQ-ID	HAIKU_UC4_HLR_7:
Requirement	The team must decrease the ATCO's subjective workload during periods of high demand.
Rationale	The assistant should help reduce ATCO's (subjective) workload (compared to the same scenario without any help from the IA).
KPA	Human Performance

5.4 UC#4 Second Validation Plan (VAL2)

Validation Objectives

The HAT objectives in UC4 have been defined to address three key Human Factors challenges for Human-AI Teaming: situation awareness, explainability and error management.

Situation awareness: The situation awareness objective requires that both ATCOs and ISA are continuously aware of the traffic situation, processing real-time data from various sources to detect trends, anomalies, potential conflicts, and anticipate future traffic patterns. Additionally, ISA should effectively communicate with the ATCO in charge by providing real-time information regarding the current sequence and, when applicable, new sequence suggestions.

Explainability: The XAI objectives require that ISA provide clear and relevant explanations regarding sequence changes. Furthermore, ISA must be able to provide different levels of explainability (CLT levels) upon ATCO's requests.

Error Management: ISA must be able to detect and inform ATCOs in charge about potentially unsafe situations, working as an additional safety net.

These HAT objectives have been aligned with the EASA HAT requirements as documented in D4.1 (HAIKU, 2024). The tables below indicate the relevant EASA HAT

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

requirements from D4.1 (HAIKU, 2024) that will be examined in VAL2, detailing the corresponding validation objective, activity, and metric.

HF Requirements

Table 68. *UC4 Cooperation/collaboration capabilities (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
HF-01: System must generate and maintain data logs on number of aircraft, departure / arrival sequences and timing, and system failures (see also HF-34).	NO			
HF-02: ISA must be able to always suggest, in real-time, the current sequence with the next three aircraft to use the runway.	NO			
HF-03: ISA must be able to always generate and show new sequences by constantly adapting to everything that is happening.	YES	To evaluate the perceived quality of the sequence generated To evaluate situational awareness, subjectively To evaluate subjective workload.	Simulator exercises with high traffic density, forcing ISA to constantly generate new sequences. Questionnaires; Semi-structured interviews (debriefing); Observation.	Qualitative: ATCO's subjective opinions
HF-05: ISA must be able to recognise a potential suboptimal sequence generated by user's interaction and suggest a better alternative to the ATCM	YES	To evaluate the new sequence generated by user's interaction subjectively To evaluate situational	Simulator exercises with high traffic density, forcing ISA to constantly generate new sequences. During the first exercise, a pilot will not comply with ATCO's instructions, forcing ISA to generate a new suboptimal sequence.	Qualitative: ATCO's subjective opinions

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

		awareness, subjectively To evaluate subjective workload.	Questionnaires; Semi-structured interviews (debriefing);and/or de-briefing Observation.	
HF-08: ISA must be able to always suggest the best possible solution to a sequence and suggest it to the ATCO, regardless of what the ATCO does	YES	To evaluate the solution generated subjectively To evaluate situational awareness, subjectively. To evaluate subjective workload.	Simulator exercises with high traffic density, forcing ISA to constantly generate new sequences. Questionnaires; Semi-structured interviews (debriefing);and/or de-briefing Observation.	Qualitative: ATCO's subjective opinions
HF-09: The ATCO must be able to easily override and turn ISA off in case of unforeseen events	NO			

Table 68. *UC4 Error Management (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
HF-28: ISA must tolerate and adjust to user manual inputs, by maintaining an ongoing sequence recommendation capability	YES	To evaluate the suggested sequences adapted to user manual inputs, subjectively To evaluate situational awareness, subjectively To evaluate subjective workload.	During the first exercise, ATCOs won't need to follow ISA's suggestions. Questionnaires; Semi-structured interviews (debriefing);and/or de-briefing Observation.	Qualitative: ATCO's subjective opinions

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

HF-29: ISA must detect and adjust to user manual inputs, by maintaining an ongoing sequence recommendation capability	NO	To evaluate the suggested sequences adapted to user manual inputs, subjectively To evaluate situational awareness, subjectively To evaluate subjective workload.	During the second exercise, ATCOs will need to follow ISA's suggestions. If not followed, ISA will detect and adjust. Questionnaires; Semi-structured interviews (debriefing);and/or de-briefing Observation.	Qualitative: ATCO's subjective opinions

Table 69. *UC4 Failure Management (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
HF-31: ISA must constantly present its functional status to the user AND System failure must result in ISA deactivation and user alert	NO			
HF-34: System must generate and maintain data logs on number of aircraft, departure / arrival sequences and timing, and system failures (see also HF-01). Users must be alerted to availability of failure log data.	NO			

XAI Requirements

Table 70. *UC4 Attributes of OpXAI (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
EXP-10: ISA must be able to generate explanations related to sequence generation (for monitoring), and ad-hoc explanations for sequence changes	YES	To evaluate the plausibility of explanations generated subjectively	Questionnaires; Semi-structured interviews (debriefing);and/or	Qualitative: ATCO's subjective opinions

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

			de-briefing including questions such as: "Were the explanations where you would expect them to be?" Observation.	
EXP-11: ISA HMI must show explanations related to sequence generation and sequence changes	YES	To evaluate the clarity of the explanations generated on the HMI	Questionnaires; Semi-structured interviews (debriefing);and/or de-briefing including questions such as: "Did explanations make sense?" Observation.	Qualitative: ATCO's subjective opinions
EXP-12: ISA must be able to show, for each sequence change, different levels of explainability with a progressive level of detail that is keyed to the expected decision / action	NO			
EXP-13: ISA HMI must provide a means for user to discover progressive levels of detail details about any provided explanation	YES	To evaluate the different explanations (different CLT levels), subjectively	Questionnaires; Semi-structured interviews (debriefing);and/or de-briefing including questions such as: "Did you notice/use all different levels of XAI?" Observation.	Qualitative: ATCO's subjective opinions

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

EXP-15: Explanations about sequence changes must be available to the user minimum delay, and permit progressive levels of detail keyed to user needs.	YES	To evaluate the explanations generated subjectively	Questionnaires; Semi-structured interviews (debriefing);and/or de-briefing including questions such as: Did those levels of XAI make sense to you? Observation.	Qualitative: ATCO's subjective opinions
EXP-16: Users must be able to access ISA explanations about sequence changes, via progressive disclosure interaction, as desired by the user.	NO			
EXP-17: ISA-generated sequences must optimise operational priorities (by default, throughput shall be optimised)	NO			
EXP-19: ISA must alert users to any potentially unsafe ISA-recommended sequences (which can be caused by outdated recommendation, or off-nominal flight trajectories).	YES	To evaluate the explanations generated, subjectively	During the second exercise, ISA will fail, suggesting a potentially unsafe situation. Questionnaires; Semi-structured interviews (debriefing);and/or de-briefing. Observations.	Qualitative: ATCO's subjective opinions
EXP-19: ISA training must cover recognition and handling of unsafe operating conditions, caused by ISA recommendations.	NO			

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

Validation Approach

Validation VAL2 will consist in evaluating key components of the assistant during several activities:

- **Overall design validation:** The assistant has been slightly redesigned following VAL1 feedback from professional Air Traffic Controllers. This new design will be tested by ATCOs in an Alicante TWR simulator. Operational relevance, integration in the simulator, acceptability, explainability of the interface and human-machine teaming will be evaluated through this activity.
- **Situational awareness:** The team must always keep situational awareness of the traffic situation, and it is key to this project to ensure that ATCOs perceive that ISA is helping them with that.
- **ATCO's workload reduction validation:** An important function of ISA consists of reducing ATCOs' subjective workload during peak hours. During VAL2 ATCOs will be interviewed after every scenario run to validate this.

Exercise Description and Scope

Table 71. *UC4 Exercise*

ID	Activity (per ATCO)	Duration	Description
A1	Briefing	20 min	ISA and HMI explanation, review objectives, etc.
A2	Exercise run #1	30 min	High traffic density, multiple sequence changes. ATCO does not necessarily need to follow ISA. A pilot does not comply with ATCO instructions at a certain moment.
A3	Debriefing #1	20 min	Interview with the ATCO to gather feedback on ISA
A4	Exercise run 2	30 min	Very high traffic density, multiple sequence changes. ATCO must follow ISA's suggestion. ISA fails at a certain moment.
A5	Debriefing #2	20 min	Interview with the ATCO to gather feedback on ISA

Find below the relation between every objective and their related VAL2 activity:

Table 71. *VAL2 activity & objects*

OBJ ID	Related VAL2 activity	Notes

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

EXP-10	A1, A3, A5	
EXP-11	A1, A3, A5	
EXP-12	A1, A3, A5	
EXP-13	A1, A3, A5	
EXP-15	A1, A3, A5	
EXP-16	A1, A3, A5	
EXP-17	A1, A3, A5	
EXP-19	A2, A4	Exercises designed to have blinking warning situations
HF-01	A1, A3, A5	
HF-02	A1, A3, A5	
HF-03	A1, A3, A5	
HF-05	A2, A4	Includes HAZOP-T0+45s-As well as (Scenario where a pilot doesn't comply)
HF-08	A1, A3, A5	
HF-09	A4	ISA turns off
HF-26	A1, A3, A5	
HF-27	A1, A3, A5	
HF-28	A2, A4	Exercise designed to have multiple sequence changes
HF-29	A2, A4	Exercise designed to provoke ATCOs the change to detect an error
HF-31	A4	HAZOP-T0-Not done: ISA doesn't work properly and check if the controllers can detect it
HF-34	Log generated after every run	

Exercise Scenarios(s)

All scenarios will be run on Skyway's Training Academy simulator.

To evaluate the usability and the explainability of the assistant (ISA), ATCOs will be briefed before starting in the simulator. During this briefing, they will be told about what to expect from ISA and the HMI explanation, review HAT objectives, etc.

Participants will then participate as ATCOs in the first run in the simulator.

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 chosen scenario is based on Alicante, with high traffic density and multiple sequence changes. ATCOs do not necessarily need to follow ISA. At a certain moment, a pilot will not comply with ATCO's instructions to see how he/she reacts to it, and to evaluate if ISA was helpful in those situations.

Afterwards, ATCOs will be interviewed as part of the debriefing process to validate HAT requirements, mainly focused on understanding if ISA was helpful or not.

The same process will be repeated again with another scenario, with very high traffic density and multiple sequence changes. ATCO must follow ISA's suggestion. ISA fails at a certain moment to test if ATCO's realise that and act accordingly.

A final debriefing will be carried out as well.

Participants and their Role

For the simulator sessions, between 5 and 10 professional ATCOs will be recruited. Gender equality will be achieved if possible.

Platform / Tools & Technique

The VAL2 simulator sessions will be performed using Skyway's Training Academy simulator.

Intelligent Assistant

Use Case 4 (UC4)'s Intelligent Sequence Assistant (ISA) is designed to support Tower Controllers at single-runway airports. The goal is to optimise runway usage by providing real-time sequence suggestions for arriving and departing aircraft. This system integrates data from simulations, estimated arrival times (ETA), and other relevant inputs to calculate and suggest the best sequence for aircraft movements.

There is an AI optimisation engine, that will be consuming the data to be collected by the different data sources and will execute the optimisation algorithm for the optimal sequence calculation, having the ability to trigger new sequence calculations based on specific stimuli, such as actions performed by the controller, new data flight data available (for example by new incoming/outcoming flights), etc.

There is also a data sink infrastructure, allowing operational and simulation systems to push data into the ISA prototype, which will be stored locally and pushed to the AI Optimisation engine for the execution of the different calculations.

Data Collection and Analysis Methods

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 qualitative data collection will be done through debriefings held after each scenario and at the end of the session. The over-the-shoulder observations will be performed by Human Factors- and Operational experts. Quantitative data will be collected through the platform's logs and recordings.

Observations: This technique mainly allows to address topics related to Human Performance, with the purpose to provide detailed and reliable information on the way the activity is carried out by the user. Direct observation enables gathering a high amount of data, especially qualitative data which cannot be collected through other methods.

In the validation exercises, direct over-the-shoulder observation will be used to collect insights about the pilot's performance, including aspects related to experienced workload, situation awareness, usability, faced difficulties, recovery actions, safety related events, etc.

Briefings and Debriefings: Briefings, debriefings and over-the-shoulder observations are interconnected techniques. This means that on the one hand, data collected through observations and briefings will be verified and discussed during the debriefings. On the other hand, insights extracted from the debriefings will be used to guide the following observations. This combination of techniques can complement and reinforce the quality of the quantitative data collected and contributes to achieving more reliable results.

System logs: System quantitative data will be collected by the extraction of log files from the simulation platform. The data will be used as performance indicators and to further enrich the qualitative information collection

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

Planned Activities

Table 72. *UC4 Planned Activities*

Activity	Activity description	General information
Overall design tests and adjustments	Tests in the simulator with several qualified ATCOs.	Madrid, Spain. September 2024-March 2025
VAL2 exercises	Validation of the solution in the simulator	Madrid, Spain. April May 2025

Table 73. *UC4 GANTT*

Legend
ConOps & Requirement Definition
Design & Development
Validation

Year	2024						2025							
Month	7	8	9	10	11	12	1	2	3	4	5	6	7	8
Human-AI Teaming concept														
AI component/s														
Non-AI technological component/s														
HMI														

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

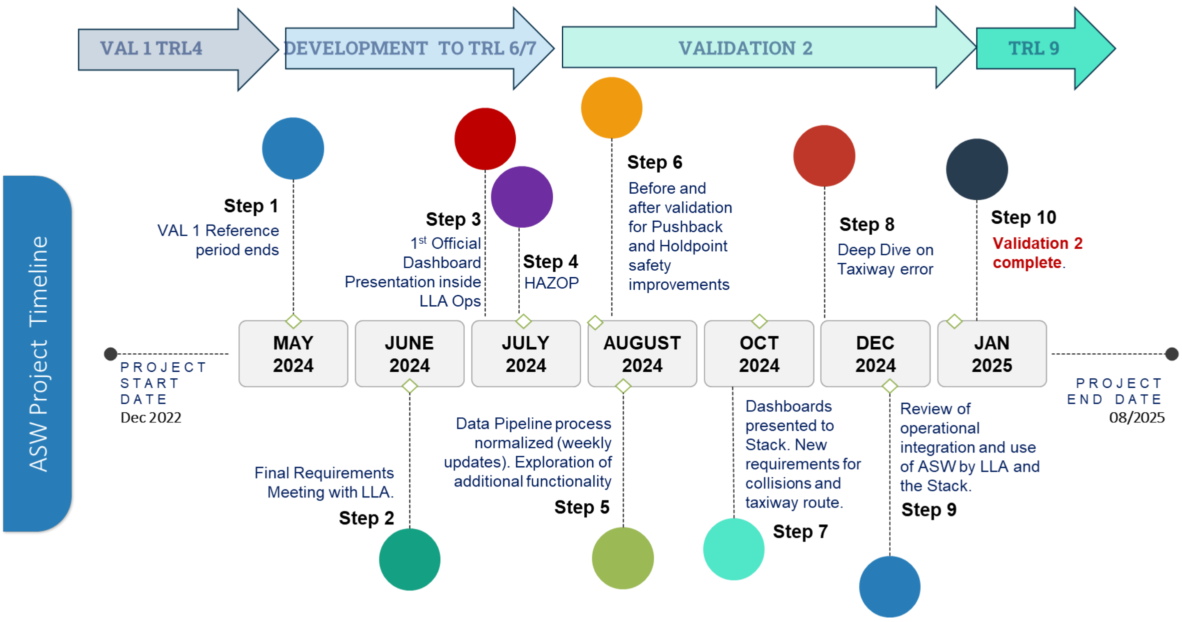
6 Use Case #5 – Airport Safety Watch

UC5 is developing an AI-powered capability to ingest and analyse a constellation of operational and safety event data collected across the airport, and to permit inferences and predictions regarding hot spots and safety 'pinch-points,' with a view to staying one step ahead on airport safety. The concept began as TRL4, but is currently (June 2024) at TRL6 and is fully expected to reach TRL9 before the end of the HAIKU project.

6.1 UC#5 TRL Overview

Figure 3 below shows the development and validation course for ASW, aimed to finish early 2025.

Figure 3. *UC5 ACW Phase 2 & 3 Development and Validation Timeline*



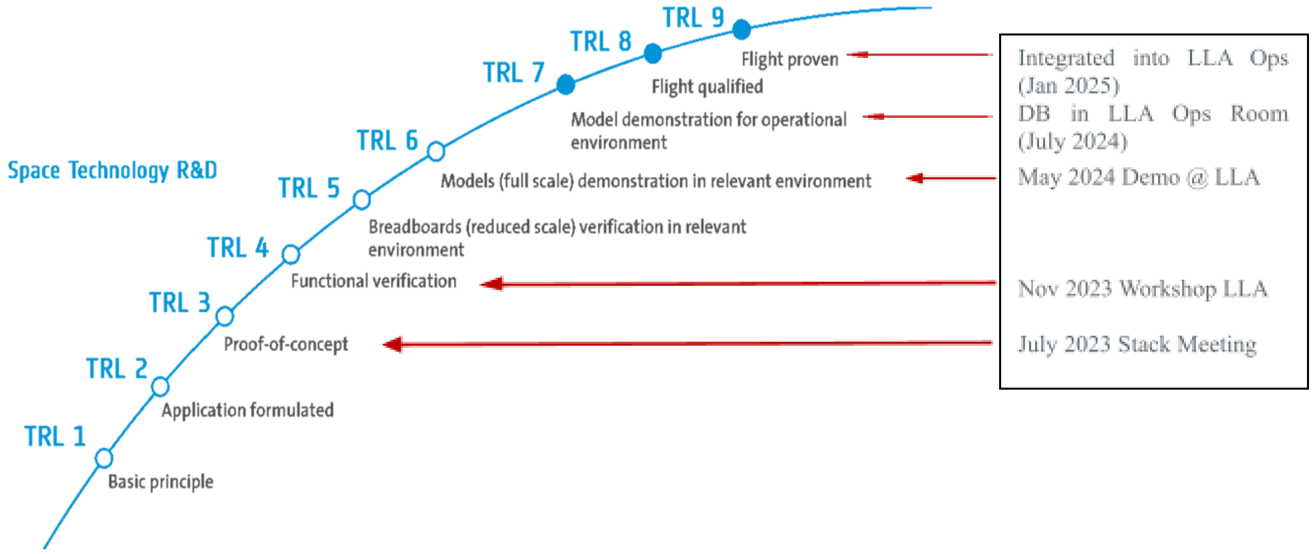
UC5 & TRL Evolution

Figure 4 below shows the evolution of the TRL of Airport Safety Watch (ASW) throughout the HAIKU project, from proof of concept during a Stack meeting in July 2023, when the Stack partners reviewed a presentation by HAIKU's UC5 team of the early Dashboard, and derived new insights for PushBack errors (TRL4), to further demonstrations and workshops verifying ASW's functionality and further requirements,

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 the ASW being deployed in the London Luton Airport (LLA) Ops Centre (August 2024) and being fully integrated into its operational safety system by January 2025.

Figure 4. *UC5 TRL Evolution for ASW*



6.2 UC#5 Validation 1 Results and Lessons Learned

In the first half of the HAIKU project timeframe, the following three aspects of Use Case 5 were validated (VAL1):

Table 74. *UC5 high level results for UC5 VAL1*

VAL1 Element	Description	Activity	Success Measures
1	Engagement of all Stack Partners in the UC5 project concept	Presentations and discussions at Stack Meetings and other ad hoc meetings	Number of Stack Meetings on UC5; Verbal agreement amongst Stack Partners that UC5 is viable and valuable; Provision of data and support by Stack Partners
2	Data sufficiency (volume and quality) for Data Science analytic purposes	Detailed analysis of heterogeneous data sets by SUITE5 / ENG	Meaningful patterns of data developed, that can be understood and interpreted by Stack Partners

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	Discovery of new safety insights	Data analysis and visualisation, and review with Stack Partners for operational safety insights	New, actionable safety insights identified for one or more of the three incident types under investigation; increased interest / provision of data from Partners

VAL1 concept results served as a proof of concept for phase 1 of the ASW research concept, which was seen as being at **TRL4**: sufficient data could be analysed in novel ways to allow operational users to identify new safety solutions to existing, long-standing safety issues (e.g. pushback error). In particular, the early results led to strong stakeholder engagement.

The subsequent challenge (phase 2) aims to ensure that such engagement is maintained while extending and refining the Dashboard (DB) user-centric requirements concerning its analytic and visualisation capabilities. The aim is also to gain further insights into holding-point busts and incorrect taxiway selection, since pushback error already has a 'win' via the safety insights generated by the Phase 1 analysis. This has resulted in three 'modes' of the Dashboard: historical, used to review across incidents; comparative, used to compare for example this month's safety statistics and factors against the corresponding month last year; and zoom-in, allowing a deep dive into an incident or a particular issue. Phase 2 brings the ASW to **TRL6.**

The final, third Phase, brings the ASW directly into London Luton Airport (LLA) operational centre (hence reaching TRL7), and concerns integrating the DB with LLA's systems, via a 'data pipeline'. This data pipeline starts with a weekly upload of data, but later on will move to real-time update as data arrives. Phase 3 also integrates the DB usage into LLA safety operations, as they occur on a daily and periodical basis. This third phase brings the ASW to TRL9.

6.3 UC#5 Updated Concept of Operations

UC5 is leveraging the ground-breaking safety efforts of LLA, including its Luton Safety Stack collaboration between intra-airport organisations (ATC, ground handling, fuel services, baggage handling, etc). LLA serves as the developmental and validation testbed for UC5. This effort was given a jump start by LLA's ongoing efforts to collect various types of safety management data. UC5 is in fact developing AI capabilities to assist on two time-scales: first, its Airport Safety Watch (ASW) IA aims to identify potential risks and issue real-time alerts. Secondly, AI-powered data mining offers the potential to analyse safety efforts and help evaluate the effectiveness of safety management initiatives moving forward.

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

Key R&D Needs and Objectives

Problem Statement – what is the problem to address with the Intelligent Assistant?

London Luton Airport is the duty-holder when it comes to safety, and as such it collects a vast amount of data from across the airport partners, creating over 50,000 entries to its safety management platform annually, all of which is categorised under its most relevant heading. The analysis of this data is undertaken manually. Right now, LLA cannot easily exploit all of this data, but with AI and Machine Learning there is the potential to identify which of our efforts produce the best results. For example, how does Incorrect Parking of Vehicles and Equipment relate to actual collisions, and what safety promotion has been undertaken to raise awareness and what was its effect? LLA has experimented with different ways of visualising safety data, including Safety Dashboards as also developed in the EU project Future Sky Safety. The problem is that such manual analysis of data, much of which represents 'weak signals', is not always timely (the analysis 'lags' behind the actual events and evolving trends), and also sometimes it is difficult to see whether the data are 'noise', or are useful precursors to potentially more serious issues.

Human Factors & Safety Risks – Potential risks and problems introduced by AI.

There should be no safety risks associated with use of the derived data analytics approach or learning tool, since it is analysing data and generating safety intelligence for the Stack, so any resultant recommendations would be evaluated by the Stack user group.

Operational Concept Description

One core concept behind UC5 is the Safety Watch IA, which provides real-time monitoring of the airport surface, identifies potential risks, and issues alerts to ATC and other concerned airport organisations. The IA also leverages historical data (e.g. runway incursion patterns) to anticipate safety breaches.

The other core concept behind UC5 is its use as a strategic safety intelligence tool. LLA currently collects more surface operational and safety event data than it can effectively process, and AI / ML seem a perfect tool to leverage in this regard. Specifically, by leveraging the strengths of human and AI teaming, applied to ML-driven big data exploration, it is hoped that improvements can be realised in safety data collection, categorization, analysis, visualisation and - ultimately - strategic decision making.

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 Targets

Table 75. *Key Performance Areas*

KPA	Category	KPI
Safety	MoP	Incident (holding point bust, incorrect pushback, incorrect taxiway selection) rates
Human Performance	MoP	Degree to which human and AI can collaborate to extract insights and warnings; Risk mitigation actions (timing and counts).

High Level Validation Requirements

The following sections summarise the HLRs and HAT requirements specified thus far for UC5. For HAT requirements (below), the associated HAT category is also presented in parentheses, based on the D3.3 Validation Framework.

Table 76. *HAIKU_UC5_HLR_1*

HL-REQ-ID	HAIKU_UC5_HLR_1:
Requirement	The team must improve the current airport safety index.
Rationale	Stakeholders main need is to manage and improve the Operational Safety at the Airport. By gaining new insights into the operational data, it is expected that the team will be able to develop new strategies that lead to the improvement of the safety index.
KPA	Safety

Table 77. *HAIKU_UC5_HLR_2*

HL-REQ-ID	HAIKU_UC5_HLR_2:
Requirement	The team must generate newer safety related insights vs. the current operational approach.
Rationale	It has been challenging to improve current safety index levels, and new insights are expected to be generated by AI-powered big data exploration.
KPA	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

6.4 UC#5 Second Validation Plan (VAL2)

ASW Concept 'Course Corrections'

Following VAL1, which can be considered to have ended by August 2023 (end of HAIKU Year 1), there have been a number of key alterations in the original ASW concept. First and foremost, there has been the realisation that a fully predictive version of ASW may not be viable until some time in the future. This is largely due to the data available. First, the incidents of interest are relatively rare, and so there is not enough data for reliable prediction. A secondary issue concerns the variable quality of the data, especially with respect to the human contribution to the event, which may be key but is often described only at a very high level or not at all. This second aspect is likely to improve as LLA itself adopts more stringent reporting templates with respect to human contributions to events, and as more data are collected. This level of data allows for the application of data science approaches, but not 'full-fledged AI algorithms'. Hence the Concept is primarily 'Advanced Analytics' rather than AI or Machine Learning, for example, though this could change as available data passes a certain threshold in the future.

A second course correction concerns the stakeholder group. Whilst there has been excellent support for, and interest in, ASW by the different Safety Stack partners, in practice very few have the opportunity to get involved with using it. Following a short Stack Partner survey, it became abundantly clear that almost all Partners wished to be informed by the ASW, rather than use it themselves or even experiment with it. Therefore, it was decided to focus on the principal and primary user, namely LLA, as they are overall accountable for safety at the airport, and have been invested in ASW since the start of the project. This course correction has made it easier to progress with the user-centred requirements, having one 'client' rather than thirty. The door is not closed definitively, however, and it is hoped for example to bring NATS back into the frame as a secondary user, since they are directly involved in the safety incidents being investigated. It may also be that as the ASW is used more, that other users (e.g. airlines, or ground handlers) begin to make requests for analyses related to their operations.

The third course correction is the move to a fully operational, integrated system. This move has been at the instigation of LLA who see the ASW as a tangible way of enhancing safety management at the airport and has of course been overwhelmingly welcomed by the HAIKU UC5 project team. This necessitates ensuring the DB systems are robust, and creating a data pipeline that is secure. As well as these technical aspects, there are organisational ones, too, concerning how the ASW will be integrated into daily and periodical safety operations. Furthermore, such a tool, aimed at safety learning, must be capable of 'learning to learn better' itself, and so cannot be considered (or designed) to be a static, unchangeable system. With this in mind, within

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 timeframe of the HAIKU project, at least two more additions are already under consideration. The first is an enhancement to the map display that is part of the Dashboard, to enable the taxiway route taken by an aircraft prior to an event, so that the aircraft's 'provenance' can be determined, as this may help better understand (and predict) future incidents. The second is the addition of a new incident category, namely collisions, referring mainly to collisions between airport vehicles and other vehicles, or between vehicles and aircraft. These are rare but can be high severity. Both of these additional capabilities, if realised, will also draw in more stakeholder engagement, the first involving NATS, and the second engaging with the ground handlers operating the various vehicles on the airport surface.

These course corrections modify or add new VAL2 requirements, as described in the next sections.

Validation Objectives

Because validation requirements in UC5 are primarily functional at a 1A level, the EASA HF and XAI requirements are not fully applicable. Additionally, UC5 is not approaching VAL2 as an experimental setup, but as a process of TRL progression. The VAL2 HAT requirements, activities and success measures for UC5 are given in the text and tables below.

HAT Requirements (HATRs), UC5

HAIKU_UC5_HAT_1 (Operational performance)
The IA must be able to process **at least seven years of historical data** on situations / incidents. Seven years is assumed to build representative past operational information for forecasting. Must verify long-term data stability, and determine how to capture data beyond this time horizon.

HAIKU_UC5_HAT_2 (Teamwork / collaboration)
Team must **identify new solutions / problems**, not previously recognized.

HAIKU_UC5_HAT_3 (Operational performance)
The IA must **generate analysis/warnings with a timeframe that support mitigation/avoidance actions**. Must still define an appropriate time horizon that supports actions. This timeframe will be calibrated on the basis of VAL1 results.

HAIKU_UC5_HAT_4 (Operational performance; Explainability / transparency)
The IA HMI must **allow what-if probing**. This will allow the team to explore and generate new insights through what-if exploration.

HAIKU_UC5_HAT_5 (Operational performance ; Situation Awareness)
The IA must be able to **process data in various formats and from various sources**.

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

HAIKU_UC5_HAT_6 (Explainability / transparency)
The IA must be able to **provide [level of] explainability regarding generated analysis**. To generate insights, understanding on the data exploration needs to be provided at different levels.

HAIKU_UC5_HAT_7 (Operational performance; Situation awareness)
The **Human User must have [level X] of knowledge on the IA analysis**. A certain level of prior knowledge must be created on the AI working, to permit critical interaction.

HAIKU_UC5_HAT_8 (Teamwork / collaboration; Operational performance)
The IA must provide for **different levels of data exploration**. Zoom in / out (both temporal and spatial) functions allow contextualization.

How UC5 HLRs and HATRs will be satisfied by VAL 1 and VAL2

The table below shows the HLRs and HATRs addressed in VAL 1 & 2 respectively.

Table 78. *UC5 HLRs and HATRs addressed in VAL1 and Val2*

Requirement ID	Short Description	VAL1	VAL2
HLR1	Improve Safety	Yes – Pushback Error	Yes – Holding point Bust & Taxiway Error
HLR2	New Insights	Yes – Pushback Error	Yes – Holding point Bust & Taxiway Error
HATR1	7 Years Data	Yes – 7yrs data ingested	
HATR2	New Solutions		Yes – Taxiway Error (Deep Dive)
HATR3	Useful Timeframe		Yes – Weekly Insights for LLA Ops
HATR4	What-if Probing		Yes – Deep Dive Exercise
HATR5	Heterogeneous Data		No – see D4.5 (multiple data sources used)
HATR6	Explainability		No – see D4.5
HATR7	User Knowledge of ASW		No – see D4.5
HATR8	Multi-level exploration		Yes – Deep Dive Exercise

How UC5 Human Factors Requirements (from EASA) may be satisfied during VAL2 exercises and activities.

In D4.1 (HAIKU, 2024), the EASA requirements relevant to UC5 are identified. Many such requirements are not relevant to ASW, since they concern explainability, spoken

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

or gestural language between human and IA, etc. Some of the requirements can be construed as potentially relevant to ASW, however, for example those in the table below:

Table 79. *UC5 EASA Human Factors Requirements*

EXP-18	Provide instructions/training to handle indications of input/output monitoring
EXP-19	Provide timely information on unsafe operating conditions
HF-01	IA situational awareness building
HF-02	User situational awareness reinforcement
HF-03	Shared situational awareness building
HF-04	Ability to submit decisions for cross-check validation
HF-06	Identify abnormal operation, share diagnosis, resolution strategy, anticipated consequences
HF-07	Detect poor decision-making by the end user in a time-critical situation
HF-26	Minimise the likelihood of design-related errors made by the end user
HF-27	Minimise the likelihood of design-related errors related to HAIRM
HF-28	Demonstrate tolerance to end user errors
HF-29	Provide opportunities to detect end user errors

These can be condensed into the following:

- Ensure safe handling and integrity of input/output data
- Ensure timely warnings (already present in Val 2 as HATR3)
- Correct (as opposed to erroneous or unsafe) situation awareness of operational safety
- Correct decision-making based on the use of the Dashboard
- Ability to detect and manage abnormal situations that may arise with the Dashboard
- Minimise design error
- Minimise impacts on the operational team and LLA safety team
- Develop tolerance to end user error (ASW resilience and robustness to end user error)

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

These issues were explored during the Hazard and operability study (HAZOP) of ASW in July 2024, since they largely concern safeguarding usage of the ASW. The HAZOP may result in further specific safety requirements for ASW's use.

Table 80. *UC5 VAL2 approach*

VAL2 Element	Description	Activity / Exercise	Success Measures
1A	Team must identify new safety insights solutions not previously recognised. In Val 1 this was done for Pushback errors.	This has since been achieved for holding point busts.	A 'before and after' validation will be made for both the Pushback Stands concerned, and the holding point where signage changes were made. Success will be indicated via either no local incidents following the change, or a reduced incident rate.
1B		Deep dive on taxiway incidents to determine new solutions.	New, actionable insights on taxiway error and associated avoidance measures.
2	The IA must generate analysis/warnings with a time frame that supports mitigation/avoidance actions.	The Dashboard will be updated on a weekly basis, and insights will be conveyed to operational LLA teams in their Monday morning briefing.	Operationally useful insights are extracted from the Dashboard and easily transmitted to the LLA airside safety teams, and can be integrated into operational engagements with other (non-LLA) ground and air staff. Feedback from LLA staff on this process is positive over several months.
3	The IA HMI must allow what-if probing	The zoom-in and causal analysis aspects of ASW will be utilised in the deep dive exercise to consider counter-factual (what-if?) hypotheses related to incorrect taxiway selection.	A record of the deep dive analysis will show how the tool can be used for such exploration and 'what-if?' Probing.
4	The IA must be able to process data in various formats and from various sources.	Establishing a stable and secure data pipeline (initially via weekly csv file transfers).	Data can be smoothly integrated on a weekly update basis without data loss or corruption.
5	The IA must be able to provide [level of] explainability regarding generated analysis	The dashboard construction follows logical analytic processes typically used by safety analysts, therefore no additional XAI is required.	The ability of the LLA safety officers to use the Dashboard without the need to ask for further explanation from the Data Scientists.

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	The Human User must have a sufficient level of knowledge on the IA analysis.	A number of sessions have been held (both face to face and remotely) concerning the usage of the Dashboard and its elements by the LLA safety officers.	The ability of LLA primary users to use the Dashboard for data analysis and to be able to glean operational safety insights, without assistance from the data scientific team, is the main operational success criterion.
7	The ASW must provide for different levels of data exploration	The Dashboard has three principal modes: historical (across incidents and other categorical classes); comparative contrasts between temporal slices of data; and zoom-in, including delving into contributory factors. It can work at macro (incident level) to micro (factor level), and can 'slice and dice' the data in numerous ways.	The full manual showing the different Dashboard panels is given in other documentation (see D4.5).
8	The ASW must be integrated into LLA operational safety approaches (new validation requirement).	The Dashboard is to be integrated into the LLA Operations Room and LLA staff break room. Insights will be given to airside operational staff prior to their operational duties, and may be included in 'points of engagement' with other airport users.	This process started in July 2024 and will be evaluated after three months to determine its operational safety utility and continued desirability. A meeting is already scheduled at LTN in early October where operational staff can be canvassed for their opinions on the weekly information arising from the Dashboard.
9	The ASW should be evaluated for potential safety risks.	A HAZOP was carried out in July 2024, considering both hazards from the way the data are collected, organised and represented (the full data preparation process and data pipeline) to how the system is used by LLA.	A successful HAZOP will identify potential hazards and mitigations to avoid risk from use of the ASW. The HAZOP will also consider predictive use of ASW, even if this functionality is unavailable at this time.

Validation Approach

The VAL2 Strategy can be broken down into three parts:

1. How UC5 serves the high-level research goals of HAIKU
2. How UC5 HLRs and HATRs will be satisfied by VAL1 and VAL2
3. How UC5 Human Factors Requirements (from EASA) may be satisfied during VAL2 exercises and activities

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

Each of these is dealt with in turn below.

How UC5 serves the high-level research goals of HAIKU
HAIKU has posed three high-level research questions, as below, with preliminary answers from the UC5 perspective.

1. What is the recommended human-AI relationship for each of the different AI aviation applications? What are the AI's purposes, the expected benefits for operations and the underlying values and design principles (e.g. cooperation, adaptability, conformity, explainability, robustness), and best means of Human-AI interaction?
    - UC5 shows how the product of humans and AI-based systems can lead to better 'safety intelligence' than either alone. The ASW Dashboard has shown that factors traditionally thought to be dominating incident occurrence (e.g. weather, traffic volume etc.), are not actually dominant; rather, it is either more local factors or the occasional confluence of certain other characteristics that make incident occurrence more likely. The Dashboard gives the human end user a broader and deeper safety landscape to explore, effectively enlarging their operational safety situation awareness. Overall, UC5 shows 'human-machine intelligence' where the human is still in control, but has augmented capabilities due to the enhanced data visualisations.

2. What does it mean for AI to be explainable? For instance, at one level the AI may operate in a way that can be grasped and understood by humans, possibly not changing the nature of the task. In more complex scenarios, the AI reasoning may not match human reasoning, given the complexity and number of variables to be considered. Is explainability a necessary precondition of trustworthiness?
    - In UC5 there is no AI where advanced explainability is required. The ASW Dashboard instead allows significantly more exploration by the human end-user.

3. How do we best teach AI, via human-in-the-loop AI learning for each of the aviation applications? Human-in-the-loop learning can lead to more explainable AI (XAI), as the user will have more visibility on the underlying mechanisms, ideally steering a continuous learning process. It can also lead to 'personalised AI', with individuals and teams developing AI assistants on the basis of their needs, abilities, limitations and experiences.
    - For UC5, the real learning that has taken place is between the operational users and the data scientists, who have come to a shared understanding (and workspace – the Dashboard) of the data and its exploration possibilities for the purposes of safety. The Data Scientists have learned what to focus on and what matters, as well as how safety works at an

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

airport. The Operational people have learned more about AI and data science, how it works and its limitations, but also the critical importance of data volume and quality.

Planned Activities

Table 81. *UC5 GANTT*

	ConOps & Requirement Definition
	Design & Development
	Validation

Year	**2024**						**2025**							
Month	7	8	9	10	11	12	1	2	3	4	5	6	7	8
Human-AI Teaming concept														
AI component/s														
Non-AI technological component/s														
HMI														

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

7 Use Case #6 – Airport Spreading Virus Prevention

The global COVID-19 pandemic demanded drastic action across the airport industry. Airports incurred additional operational expenses for extra cleaning and sanitization along with touchless/frictionless travel. Public health services in every country released guidelines and best practices for the persons that should be present in an indoor space and the distance between the persons. To be able to fastly and effectively face potential future pandemics, solutions for preventing virus spreading in crowded places, such as airports, must be studied and designed. Technological support to automatically routing passengers in airports could be essential to support the aviation industry in facing potential future health challenges related to pandemics.

7.1 UC#6 TRL Overview

Table 82. *UC6 TRL Overview*

COMPONENT: COVAID tool of Android application with AI model and hardware prototypes for person queues and indoor air quality sensor board and system.		
TRL	**Month**	**Activity to reach selected level**
1	1	Literature Review and initial theoretical formulation
2	9	Theoretical foundation of solution and initial implementation of subsystems
3	16	VAL1 with initial prototype. Injected data and very basic information from staff members regarding implemented subsystems
4	27	Finalised prototypes and subsystems. VAL2 at the Amygdaleona airport in Kavala Greece with multiple passengers
5	36	Acquisition of more data. Fine tuning of solution to reach TRL5
6	-	

UC6 is targeting TRL5. The main factor inhibiting the achievement of TRL6 within the HAIKU project timeframe is:

- The airport where validations will take place. It is a General Aviation Aerodrome and not a commercial one.

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

7.2 UC#6 Validation 1 Results and Lessons Learned

Lesson 1: Wireless camera systems showed robust performance during testing, though environmental factors such as interference and signal range must be accounted for in future iterations. This testing highlighted the importance of optimising both hardware and software for reliability in real-world conditions.

Lesson 2: The API calls between the cameras and the air quality sensor functioned as expected, but occasional delays in data transmission were noted. Streamlining communication protocols can improve efficiency and reduce latency for real-time monitoring.

Lesson 3: The infection index classifier delivered promising results in detecting relevant patterns. However, further tuning and additional training data is necessary to improve accuracy and adaptability across varied conditions.

Lesson 4: Simulating multiple passengers successfully mimicked real-world usage patterns for database insertion. However, database performance may degrade under high-load conditions, warranting optimization for scalability.

Lesson 5: The selected classifier for air quality forecasting performed satisfactorily, though it occasionally required human intervention to ensure accurate forecasting. Further fine-tuning is required to reduce the dependency on human input.

Lesson 6: The chatbot for passenger communication and explainability of routing recommendations was effective in real-time interactions, though user feedback suggests that enhancing the clarity of explanations and responses can improve user engagement and satisfaction.

Lesson 7: The weighting factor used to recommend optimal routing performed well in simulations, balancing multiple variables effectively. However, some adjustments are necessary to account for edge cases where congestion or delays might not be immediately predictable.

Lesson 8: Distance and queuing data provided valuable inputs for calculating the optimal route. However, real-time queuing dynamics, especially in dense environments, presented challenges in data accuracy, necessitating more sophisticated sensing or filtering techniques.

Comprehensive white-box, black-box, and unit testing across subsystems validated the core functionality of each component. These tests confirmed that each subsystem met its design specifications. The integration of all subsystems was tested with injected traffic, confirming that the system operates cohesively under controlled load conditions. While some minor issues were identified, the overall system was functional.

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

Extended validation with real traffic demonstrated that the system maintains its integrity and performance in real-world conditions. Key performance indicators, such as response time, classification accuracy, and API call efficiency, were within acceptable limits.

7.3 UC#6 Updated Concept of Operations

Key R&D Needs and Objectives

Problem Statement – what is the problem to address with AI?

Travel and passenger mobility present a considerable risk of COVID-19 infection to passengers, due to confined space and recirculating air (both in terminal and aircraft). The object of our platform is to provide real-time information about the indoor conditions facilitating the spread of contagious diseases such as COVID-19. Our tool is a data-centric platform based on a wireless network of Internet of Things (IoT) sensors strategically installed in indoor environments to facilitate the monitoring and prediction of the critical risk factors associated with the spread of COVID-19. IoT devices are equipped with an ecosystem of sensors such as LiDAR to measure the distance between passengers and an assembly of humidity, CO2, and tVOC (total Volatile Organic Compounds) sensors measuring indoor air quality. Dedicated AI/ML inference models are used to forecast the arrival rates at the boarding points and the evolution of the air quality-related measures, in real-time. These arrival rate models are used to model passenger flows and operations in the terminal to build risk-free "what if" scenarios of how infections can spread and what potential measures can be taken to better manage infection risk.

Essentially, the proposal is a mobile phone recommendation system, which takes the camera sensors as input to provide the occupancy and the number of passengers moving towards the airport common places. This system substantially minimises the occupancy and queues in every common place, thus, preventing the places from being overcrowded. The lidar sensor obtains the distance between the passengers in the waiting areas and while moving. Thereafter, the air quality is monitored to see whether intervention in the ventilation is essential.

Human Factors & Safety Risks – Potential risks and problems introduced by AI.

Problems in data collection lead to false alerts. Dynamic nature of information flow due to passenger mobile phone application, leading to real-time change in the occupied space. Passengers do not follow the routing recommendation of the system, leading to AI to be inconsistent.

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

R&D Objectives

Table 83. *UC6 Key R&D Objectives*

OBJ-ID	Validation Objective	Success Criteria ID	Success Criteria
UC6-OBJ-01	To assess the acceptability and operational efficiency of the recommendation by the user using their mobile phones.	UC1-CTR-01	The recommendation is considered to be operationally efficient and acceptable in terms of the routing recommendation.
UC6-OBJ-02	To check the interconnection between sensors, application and other subsystems as an integrated one.	UC1-CTR-02	The establishment of coherent and punctual operation and cooperation of all subsystems.
UC6-OBJ-03	The operational efficiency of the health and safety system which will be operated by the health and safety staff.	UC1-CTR-03	Acceptance of the solution and verification of applicability of the outcome to the requirement.
UC6-OBJ-04	Black box and white box testing of all subsystems comprising the complete system.	UC6-CTR-04	Verification and testing success.
UC6-OBJ-05	Investigate passenger feelings through statistical analysis with data obtained by questionnaire to assess the system.	UC6_CTR-05	Validation of user-centric system performance

Operational Concept Description

The recommendation AI system is an integrated IoT-based system which will feed near real-time data to the passengers regarding the best possible route inside the airport's common spaces. The sensors that are used are cameras monitoring occupancy and queues, and real-time counting of passengers, as well as lidar to measure the distance and air quality sensors to monitor the air in the indoor environment. The recommendation system obtains the infection probability (high/low) as a categorical variable that is the source of the ML model that runs on the cloud. Thereafter with the POST method the recommendation is fed to the passenger's mobile phone with the appropriate routing set using a weighting factor comprising the occupancy, difference in queues and people moving towards each common space. Note that the preferences of the passenger are set by the mobile phone application and are stored in the respective database field, which will be populated according to time constraint to a different field (e.g. 2 minutes). A chatbot engages the passenger to Q&As whereby a

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

knowledge base has been populated to check for the closest question to provide the respective answer. Moreover, the chatbot automatically provides the CLT levels and it can get values from the sensors by querying the database. Specific questions regarding the sensor values will be handled in the near future by querying the database as well.

The system ensures bi-directionality since the recommendation part of the system essentially resembles a chat application with the AI bot being one person the passenger being the second. Different stages of explainability are introduced in this manner, depending on the time constraints identified. The system can have a conversation with the passenger in different time frames. Focusing on the AI responses, the system recognizes questions based on similarity to respond accordingly to the potential clarification needed by the passenger.

Finally, the air quality of the indoor spaces will be monitored, such as tVOC, CO_2, temperature and humidity, in order to correlate them with the ventilation and assess the current settings.

Performance Targets

In UC6 the primary KPAs are the quality of recommendation, the infection suggestion output, the routing output and the air quality output. The KPIs include but are not restricted to the routing outcome, monitoring of passengers, the distance between them, the queue of each common place, the air quality forecasting index. Following this, the respective metrics that are identified are, sequence of places to visit, the passengers' occupancy, number of persons moving towards each common space, square metres per person, the number of persons in the queue, and air quality.

Table 84. *UC6 Performance Targets*

KPA	Category	KPI
Optimal routing	MoE	Effective calculation of Capacity, Queues and Incoming Passengers to be fed to IA
Obtaining perfect information about occupancy	MoP	Persons counting in/out from sensors
Obtaining air quality information	MoE	Air quality metrics to be obtained by sensors
First response available promptly to Passenger	MoP	Speed of IA calculation

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

Occupancy to be populated every two minutes	Constraint	Counting of persons
Passenger phone should immediately send preferences to populate database	Constraint	Connection to server, injection of data
IA bidirectional communication and learning	MoE	Chatting with IA chatbot, send data, obtain response
Classification of air quality should be available promptly to Health and Safety Officer	MoP	Speed of classification model
Effective calculation of Capacity to be fed to IA	MoE	Persons per time

High Level Validation Requirements

Table 85. *UC6-HLR-01*

HL-REQ-ID	UC6-HLR-01
Requirement	The IA needs to provide optimal routing of passenger
Rationale	This will make sure that the IA will send the perfect information to the Passenger based on the input data
KPA	Optimal routing, IA performance

Table 86. *UC6-HLR-02*

HL-REQ-ID	UC6-HLR-02
Requirement	The IA and passenger needs to obtain perfect information regarding the occupancy, queues and incoming persons in common places
Rationale	The hardware needs to provide accurate results in a timely manner. The IA will obtain these values and provide them to the passenger in the form of recommendations and answers to questions
KPA	Obtaining perfect information about occupancy, queues, incoming data. System performance

Table 87. *UC6-HLR-03*

HL-REQ-ID	UC6-HLR-03
Requirement	The IA needs to obtain air quality measures to be available for the correlation and classification
Rationale	This ensures the communication between the IA and the air quality prototypes and the perfect information that the IA needs to obtain

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

KPA	Obtaining air quality information, System Performance

Table 88. *UC6-HLR-04*

HL-REQ-ID	UC6-HLR-04
Requirement	The first response should be available within 30 seconds of the evaluation process
Rationale	The passenger needs to be engaged in the HAT process as soon as possible by the IA. The current system is quite fast, but it will be tested with high volumes of passengers.
KPA	First response available promptly to Passenger, System Performance

Table 89. *UC6-HLR-05*

HL-REQ-ID	UC6-HLR-05
Requirement	The occupancy, and queues of the common places should be populated every 2 minutes in the database
Rationale	This constraint need to be satisfied in order to get near real-time information which makes sense with the time the Passenger needs to move to certain common places of the airport
KPA	Occupancy and queues to be populated every two minutes, sensor and system constraint

Table 90. *UC6-HLR-06*

HL-REQ-ID	UC6-HLR-06
Requirement	The passenger phone should immediately send the preferences to populate the database
Rationale	This ensures that the information set to the phone of each Passenger, in the form of preferences, gets immediately to the database via the web service.
KPA	Passenger phone should immediately send preferences to populate database, Constraint to be satisfied

Table 91. *UC6-HLR-07*

HL-REQ-ID	UC6-HLR-07
Requirement	The IA needs to communicate bi-directionally and learn from the passenger responses
Rationale	This is the heart of the IA which will provide real time explanation to the passenger through the chatbot.

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

KPA	IA bi-directional communication and learning, System effectiveness

Table 92. *UC6-HLR-08*

HL-REQ-ID	UC6-HLR-08
Requirement	The classification of the air quality should be correlated and be available to the Health and Safety operators in less than 5 minutes
Rationale	Information is essential to the operator to correlate the microclimate with the occupancy of common places.
KPA	Classification of air quality should be available promptly to Health and Safety Officer, System Performance

Table 93. *UC6-HLR-09*

HL-REQ-ID	UC6-HLR-09
Requirement	The capacity and queues as well as the expected persons heading to each place of the common places of the airport need to be effectively calculated to be fed to the IA
Rationale	People do not like to get overcrowded not only for COVID but for waiting purposes as well.
KPA	Effective calculation of Capacity to be fed to IA

7.4 UC#6 Second Validation Plan (VAL2)

Validation Objectives

HF Requirements

Table 94. *UC6 Cooperation/collaboration capabilities (EASA Macroarea)*

EASA-like requirement (D4.1)	**Covered**	**Validation Objective**	**Activity**	**Metric**
HF-01: System must keep track of environmental events, user objectives, crowd patterns, user preferences, routing and sequence recommendations, and user non-compliance events..	YES (without the user compliance, since it is done indirectly)	To validate routing sequence, likelihood of infection and environmental events based on sensor and mobile phone data	We address these in the mobile phone application and the online air quality tool. Introducing cases of crowded places in the airport.	Occupancy, queue, mobile phone preferences to feed the weighting factor and ML algorithm. Also, air quality metrics 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

				environmental sensor data.
HF-02: System must be able to populate and update its database of input conditions and associated recommendations.	YES (associated recommendations are not stored)	To make sure that the database is properly populated and the data stored based on experimental subjects participating in experiments.	We check the database insertion and cross-reference it with the data from the sensors and subjects participating.	People going in this direction, queue and occupancy metrics. Also, csv values of the air quality.
HF-02: System must be able to provide ranked recommendations.	YES	Ranking based on majority of passenger comments. For example the majority of passengers commenting if a toilet is busy will appear as a result of a following question by another passenger.	Check ranking based on person counting.	Ranking column of csv.
HF-03: The system learns from passenger (group) history, and adapts its recommendations over time.	PARTLY	The objective is to assess the metrics used in the routing algorithm and the likelihood of infection based on sensor data and mobile phone preferences. Hence the passengers contribute to the history of data used directly and indirectly.	We place the environmental values to the csv to retrain where necessary. We Get the latest database field values to calculate the routing sequence and the NLP based on the ranking and the time elapsed.	Cross-reference with tools measuring data. Check data to cross-examine the routing sequence and the answer from the NLP.

Table 95. *UC6 Error Management (EASA Macroarea)*

EASA-like requirement (D4.1)	Covered	Validation Objective	Activity	Metric
HF-28: The system provides updated routing recommendations and course / sequence corrections,	YES	To engage the passenger to the routing process to minimise likelihood of infection by	Make the passenger engaged in a discussion with the AI regarding the AI decision and CLT	Rerouting by securing the database fields are populated and persons measurement

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

following passenger error / routing non-compliance		getting involved in the NLP discussion as well.	levels will strengthen that.	reflect the real cases (by experiment)

XAI Requirements

Table 96. *UC6 Attributes of OpXAI (EASA Macroarea)*

EASA-like requirement (D4.1)	**Covered**	**Validation Objective**	**Activity**	**Metric**
EXP-12: System must be able to explain its recommendations (regarding passenger route and sequence changes) at different levels of explainability, and with progressive levels of detail	YES	To provide the evidence to the passenger regarding the routing sequence with data coming from the cloud AI model.	We provide the passenger with data that comes from the sensors and mobile phones and present them to understand the procedure	CLT levels with sensor and weighting factor data.
EXP-13: System must be able to explain its passenger routing recommendations at different levels of abstraction, depending on both task and user properties	YES	To provide different levels of explainability based on the CLT levels	Provide different values coming from the database	Database fields.
EXP-14: The system must be able to provide passenger the ability to customise levels of explainability associated with any provided recommendations	NO	The CLT levels provide relative information to passengers, based on different values over time	We query the database and get values that may be different for CLT levels	Database fields
EXP-15: The system must be able to explain its routing recommendations (at progressive levels of detail) in a timely manner, and tailored to the user and operational properties	YES	The CLT levels are designed to provide information regarding the routing process in a timely manner (using timers) and current database values	Provide the passenger with information based on the sensors and mobile phone insertion.	Database fields queries and cross-examine with setting of experiment.

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

EXP-17: The system must provide the passenger valid and reliable explanations, consistent with information obtained from sensors and database fields	YES	This comes as a result of the functionality of the sensors	The database fields are presented to the passenger	Data coming from the database (current last row)
EXP-19: The system must inform the passenger of any error modes, and unsafe AI based operating conditions, to make passenger aware of possibly unsafe operating conditions	YES	To provide the system with faults of communication and connectivity with the cloud, which influences the operability	To directly monitor the functionality of the application from device to the cloud.	Connectivity, database retrieval, csv lines.
EXP-19: The system must inform the passenger of any error modes, and unsafe AI based operating conditions, to make passengers aware of possibly unsafe operating conditions. ** Training must cover potential unsafe operating scenarios	NO	The system does not cover error modes since it is dynamic in a sense that if a number of passengers do not adhere to the recommendations, crowded scenarios will occur.	The chat module assists to raising the awareness of these scenarios pointing out personal responsibility	Chat Q&As with data from sensors.

The following motivations for respective EASA HF and EXP objectives are given.

HATR1: Relevant Explainability

The chatbot functionality will be validated with respect to the questions or comments inputted by the passenger at the Amygdaleona airport in Kavala. The CLT levels will be investigated in a practical manner according to the responses given by the system. The SHAP method for the air quality tool will be investigated based on the newly formed dataset.

HATR2: Abstraction of Explanations

The explanations given by the notification tab and the chatbot respectively, will be validated to check their optimality and correctness. Abstraction will be given based on the general questions to the chat, while specific responses will be validated coming from the sensors as explanations of the system.

HATR3: Operational Explainability Customisation:

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

Passengers can personalise their interactions with the chat by asking questions of varying specificity. For example, a passenger might inquire if the duty-free shop is generally crowded or if the restroom is currently busy. The chat will respond based on previous user feedback for the first question and by accessing real-time sensor data for the second. The responses will be validated against the values from the sensors.

HATR4: Timing of Explainability

The CLT levels used for explainability have set time values for different explanation levels. The chat and initial routing sequence follow these levels. When a passenger asks about the overcrowding of a specific place, the system retrieves the necessary sensor data to address the question, considering the time-critical nature of the query. The timing will be validated as the time constraints come from the definition of the CLT paper.

HATR5: Validity of Explanation

The system's explanations will remain valid as long as the sensor data is accurate, ensured by a fully integrated system. In the chat Q&A, responses will be filtered and ranked to prevent irrational answers. The NLP model will generate the best-matched response to the passenger's query.

HATR6: Unsafe-based System Operation

The passenger will evaluate the AI model's decisions by visiting common areas in sequence, assessing the decision's optimality. If the passenger believes the AI made an error, she can recalculate the route to check for changes and leave a comment in the chat. Misuse of the application is unlikely, as it relies on correct use by the passenger. The validation of unsafe operation will be undertaken by designing specific scenarios of different types of passengers.

HATR7: AI-based system self-situation representation

The AI-based system gathers data from camera sensors and passenger preferences, making the output dynamic and passenger-influenced. The NLP model adapts to passenger Q&A, ensuring the system operates autonomously and responsively. These dynamics will be captured and assessed at the airport with different passenger scenarios.

HATR8: AI-model individual situation awareness enforcement

In UC6, the system can only offer recommendations. The system aims to raise passenger awareness, which can sometimes act as enforcement. However, it ultimately depends on the passenger to follow the instructions and 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

Validation of individual situation awareness may come from the ranking of the Q&As which will ultimately change the responses of the chatbot.

HATR9: Shared Situational Awareness of AI-model

This system relies on passenger input to dynamically update its knowledge base, altering situational awareness and responses accordingly. Unlike traditional NLP with predefined answers, it adapts, changing both the passenger's perspective and the AI's responses. The AI's situational awareness will be validated as above while the situational awareness of the passenger will be assessed by providing them a questionnaire on the user experience with the COVAID.

HATR10: AI-based model Adaptivity to Abnormal Operations.

The COVAID system is advisory only, meaning that the passenger can challenge the AI's routing process. The AI responds using sensor data and input from other passengers to justify its recommendations. If a problem is identified, the AI may suggest re-routing to verify the issue. In complex situations, the AI can compare common area indexes to aid communication. Again, abnormal behaviour will be assessed with real passengers at the airport.

HATR11: AI-based system task allocation pattern acceptance

The system is dynamic and adapts to passenger movements and routes within common airport areas. Weighting factors vary based on the type of location, influencing occupancy times. Task allocation patterns are accepted based on the movements of multiple passengers, not solely the application user. Overall, it is a system that evolves continuously over time. The time evolution will be validated at the airport premises with the passengers using COVAID.

HATR12: Spoken Language Utilisation

The IA in UC6 is currently not intended for spoken language processing, but there is consideration for implementing it in the chat, depending on time constraints. Language can be effectively used in the chat, and responses will be provided in the same format as written text. In the case of implementation, the responses of speech-to-text and vice versa will be validated by passengers.

HATR13: Spoken Language Misunderstanding

The AI should utilise an algorithm to accurately process spoken language input and respond effectively. It must filter out various types of voices and background noise, while addressing potential misunderstandings through Q&A interactions. The algorithm will be validated by using different passenger voices and accents.

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

Validation Approach

The validation exercise will have to assess the different subsystems that comprise the full system. Essentially, the KPAs and KPIs will be evaluated for the operability of the subsystems and the service they offer to the passengers and health and safety officers. The second validation will consist of the interoperability of the components of the system and elaborate on the integration at the first level with a number of passengers coming from the Amygdaleona airport in Kavala, Greece. As for the AI engine the exercises that will be performed will consist of the response evaluation based on the user input and the explainability coming from the bi-directional communication.

To answer the question "did we build the right team", the chatbot application that will provide the answers to the passenger questions provides exactly that. The Passenger query the knowledge base where the IA will get the best response. The ranking of the similar questions will be assessed by updating the respective field in the knowledge base if a passenger selects a similar answer. Also, the date and time will be changed every 10 minutes to offer dynamicity to the application, and the ranking will be reset. In the opposite direction, the IA will provide the necessary information coming from other passengers and from the sensors to the Passenger via the NLP chatbot, in order to convince her regarding the optimality of the recommendation. Certainly, the two agents will team-up to get the optimal routing.

In terms of if we are building the right thing or if we have the right design, for the former, the right specification can actually answer in conjunction to the operational requirements identified. For the latter the right design in any system is crucial and this leads to the correct implementation of the system. There are methodologies in the literature that allow a detailed design. Of course, amendments usually take place during the system implementation.

Exercise Description and Scope

The validation exercise will commence with the input of preferences by users present in the Amydgaleona airport in Kavala, Greece, using their mobile phone application. Thereafter, the AI intelligent assistant will output the appropriate likelihood of infection and routing sequence to the notification page of the application, in the respective time constraints. This constitutes the primary requirement of the UC since the routing of the passengers will take place using this validation exercise. The primary objective is the optimal selection of common places based on the weighting factor of each place. This will be the primary exercise. Another important exercise is the validation of the NLP answers to the passenger questions. Further, the CLT levels automatic response at the chat will be validated with respect to data coming from the weighting factor and the sensors. Smaller exercises will be done for the evaluation of the subsystems, 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

the sensor output, the database storage as well as the health and safety portal with the respective IA. Finally the statistics coming from the database will be evaluated.

Table 97. *UC6 Exercise Description and Scope*

Exercise Description	Scope (What is being assessed)	Key R&D Objectives explored
Occupancy Measurement	Person counting	Person detection and counting.
Web service connection	Information obtained from phones to database	Kotlin to php and postgresql connection
ML model accuracy	Initial Random Forest classifier of infection index	Accuracy, error of classification
Chat application and explainability	Bot communication with Passenger	Knowledge base population, question similarity and closest match, update of ranking and date time, explainability based on information obtained from sensors based on CLT
Air quality sensor accurate readings	Collection of data and transmission to server and database	Hardware and software implementation
Air quality classification and human engagement	Classification and HITL	HITL
Queuing of persons counting	Distance and queue from prototype	Tracker and lidar operations
Passenger incoming to common place	Passenger preferences taken from mobile phone	Update the respective database field on the cloud.

Exercise Scenarios(s)

The primary scenario is the collection of the present users at the Amygdaleona airport in Kavala Greece who will use the COVAID system. Essentially, each user will place her preferences on the initial tab of the mobile phone application, and then she will have to move to the second tab to check the infection likelihood and the routing sequence. Thereafter, the system will provide explanations to engage the passenger to a conversation in the chat. Here CLT1-5 levels will be used with CLT1 being the output in the second tab. The passenger will also have the opportunity to engage in Q & A

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

activities in the chat and again, using the appropriate questions data from the sensors will be displayed.

When the user visits one of the places from the routing recommendation she can ask for a reroute to check if the situation with overcrowding has changed. Then the system will engage in the same procedure as well. This scenario will comprise the validation of operation of all the subsystems as described in the exercise description and scope.

In terms of health and safety and air quality portals, the database will provide the statistics and a system whereby the operator will be able to check the classification will be given. As such the air quality's classification validity from the ML model will be provided. The statistics will be given to the operator.

The reference scenario for the routing system will comprise the passenger injection of data to the database from residents of the airport, and the responses from the user to the system chatbot to check bidirectional communication.

For health and safety, the statistics will be provided using data coming from the sensors deployed at the airport. For the air quality again real data will be provided whereby the output will be a priori known.

Participants and their Role

The participants of VAL2 are mainly the candidate pilots of the Amygdaleona Airport in Kavala Greece, who will act as passengers. The COVAID tool will be given to them to assess it and validate its operation and efficiency. Moreover, the health and safety portal with the air quality information, classification and XAI will be assessed by the staff of the same airport.

Platform / Tools & Technique

The validation tools are a mobile phone application, a portal, a ML model UI and a NLP chatbot. The equipment used are commercial cameras, a wireless camera prototype, air quality sensors, a lidar sensor and the respective programming languages, such as python, Java, PHP and others that may emerge.

The technique of the validation in terms of the wireless camera prototype, comprises the comparison of different trackers, i.e DeepSort, to distinguish the different IDs of different persons within the view of the camera. Moreover the new algorithm that has been developed will be put to the test for tracking purposes. Note that the wireless camera prototype will be used for person detection in order to obtain the queue within the view. This algorithm essentially does not use a known tracker, but it obtains a unique ID per person by checking the sequence which she comes into the view. This is as a result of the difference of frame tracker IDs when a video and a live feed is

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

utilised. Queuing is managed and when it comes to distance a lidar is tested to get it from the pcap file of the measurement.

In terms of the application we fed data to the web service via the android application and saw that the injection is successful. For the second validation, and since the installations of the commercial cameras have been undertaken, real data coming from the sensors will be fed to the web service and the database there. The wireless camera prototypes will also be checked for their valid responses according to their views. For the chat- application, the NLP model has been initially tested in the laboratory for its validity. However, it will be further investigated and validated at the airport as part of the complete system.

In terms of the infection classification and routing, we selected the random forest with artificial data we constructed and we aim to test the weighting factor to route the persons to the less crowded places. The second validation will aim to the collection of a dataset and proper training using real data.

For the air quality statistics and classification, different classifiers will be constructed and an interface will be available to the operators. Here a dataset will be formed based on the sensor readings and the classifiers will be put to the test to check their results and accuracy.

Intelligent Assistant

The IA of the COVAID primarily provides the routing sequence and the likelihood of infection based on the weighting factor with metrics from the sensors deployed and the classifier respectively. The routing sequence is the primary concern; however due to the fact that it does not provide information that enhances the trust between the COVAID and the passenger, the AI chatbot service is available to answer potential questions of the passenger relevant to the scenario and give sensor data and weights that explain the routing decision, in order to make the passenger more comfortable with the technology. This will provide the passenger with the means to participate and feel the responsibility of the routing decision. Moreover, the web portal of the statistics, classification and XAI for the health and safety officers offers an assistance towards understanding potential problems in the air quality at the airport.

Data Collection and Analysis Methods

The data collection will be primarily done using a PostgreSQL database whereby the necessary tables and fields will be created in order to articulate on the methods used. In particular, the routing table has fields that determine the people coming towards each common place taken from the preferences, the occupancy of each common place coming from the camera sensors as well as the queues and the respective

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

distance. Also data collection can be done in the chatbot but the person making a comment and the knowledge base to transform it to a Q&A, if the similarity between the available responses and the one given by the passenger is low.The majority of the fields are quantitative. For the statistics table the lot of the values will be utilised. In terms of the air quality data, the best option is to store the data to a table whereby the classification will be done using the appropriate tool implementation.

The analysis method for the routing application will be the assessment of the weighting factor as well as the different bidirectional communication strings in the chat that will be used and recognised by the system. The weighting factor is a method coming from stochastic network optimization and it has been successfully used in wireless networking research. The only concern we have regarding the usability of the approach is the fact that the queues are First-In First-Out (FIFO) and not Last-In First-Out (LIFO), which is the optimal.

Notably, for the lidar, the distance will be obtained by assessing the pcap file of the measurement together with the json file of the configuration.

For the health and safety portal, standard statistical analysis will be provided. For the air quality. different classifiers will be used to get the best accuracy. Statistical correlation between the person occupancy and the trends of person movement will be given to the operator, who will be able to ask questions or intervene regarding the importance.

In terms of explainability, a Shapley values (Shapley, 1953) method is underway which is mapped to a potential game (Monderer & Shapley, 1996) approach, which when proven almost converges to a Nash equilibrium (Nash, 1950). If we can show that then the Shapley values will be shown with a single function, the potential function. The price of the stability of the optimisation will be shown to be 1 (on a 0-1 scale), meaning that we may accomplish the social optimum. In terms of the SHAP method (Lundberg & Lee, 2017), investigation is done using an implementation to distinguish the Shapley values computed and how we can apply our 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

Planned Activities

Table 98. *UC6 Planned Activities*

Activity	Activity Description	General Information
Prototype testing	Testing of the wireless camera prototype for queues	Testing of view queues at the airport with real people.
Networking testing	Testing of the API calls of the cameras and the air quality sensor	Testing the effective operation of the devices and their connection to the server.
Infection ML classifier	Testing the classifier used for the infection index	Make a dataset and calculate the index which will be validated.
Multiple sources of database insertion	Insert data into the database mimicking multiple passengers	Check with the number of passengers
Air quality classifier and statistical tool	Testing the classifier selected for the air quality forecasting.	Making a dataset to be used.
Chatbot and passenger communication	Testing the application Q&As and the explainability of the application with respect to sensor readings	Use the airport persons
Weighting factor evaluation	Testing the weighting factor to recommend optimal routing	Small scale at the airport
Queuing and distance measurement	Testing the distance and the queue of people to feed the weighting factor	Demo produced with prototype and lidar
Subsystem verification on site (VAL2)	1, 2, 3,5, 7, white box, black box and unit testing	Q3 2024
Overall initial validation (VAL2)	initial validation of 1-8, with real data at the airport	Q4 2023 – Q1 2025
Extended validation (VAL2)	extended validation of 1-8 with real traffic	Q2 2025 – Q3 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

Table 99. *UC6 GANTT*

	ConOps & Requirement Definition
	Design & Development
	Validation

Year	**2024**						**2025**							
Month	7	8	9	10	11	12	1	2	3	4	5	6	7	8
Human-AI Teaming concept														
AI component/s														
Non-AI technological component/s														
HMI														

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

Conclusions

In reflecting on our efforts to harmonise work across six diverse use cases, several lessons were drawn:

- **Diverse Nature of Use Cases**: The six use cases varied significantly in both content and the nature of the partners involved. This diversity made harmonisation challenging, particularly as industry and research partners approached the task from different perspectives. The harmonisation was pursued to the extent that it was beneficial - to ensure that the overall research questions were being addressed - avoiding to overfit the different Use Cases into one unique format.

- **Industry vs. Research Perspectives**: Industrial partners tend to emphasise detailed, design-focused requirements, primarily aimed at communicating with regulators. In contrast, research partners are more interested in knowledge-based requirements, focusing on understanding how specific systems function and exploring different design options. This difference in focus constitutes a communication challenge.

- **EASA Framework**: We employed the EASA Framework to harmonise Human-AI Teaming and safety requirements across use cases. While this helped in creating a common background and aligning requirements, the framework primarily addresses Human-AI Teaming and safety and does not capture the broader research complexities. Thus, it proved useful but insufficient for covering all aspects of the use cases.

- **Team Composition**: A key recommendation for future projects is to form “the right mix” teams for each use case, comprising both industry and academia. This would include a problem owner from industry, an AI provider (industry or academia), and a human factors expert (industry or academia).

- **Balancing Harmonization and Diversity**: While harmonising the approach adds value by promoting alignment, it is important to acknowledge that the diversity of the use cases inherently limits full harmonisation. Each use case has unique challenges that may not fit neatly into a single framework or 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

References

BEA. (2012). Rapport final, accident survenu le 1er Juin 2009, vol AF447 Rio de Janeiro—Paris. https://bea.aero/docspa/2009/f-cp090601/pdf/f-cp090601.pdf

EASA (2024). EASA Artificial Intelligence (AI) Concept Paper Issue 2: Guidance for Level 1&2 machine learning applications, March 2024. Online at: https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-issue-2

EASA (2023). EASA Concept Paper: First usable guidance for Level 1 & 2 machine learning applications (proposed Issue 2), released 24 Feb 2023 (https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-concept-paper-proposed-issue-2-open). European Union Aviation Safety Agency.

EASA (2020). Artificial Intelligence Roadmap: A Human-centric Approach to AI in Aviation. European Union Aviation Safety Agency. Version 2.0. February 2020.

HAIKU (2024). Deliverable 4.1: Operational System/Technical Requirements. Version 1.0. September 2024. Human AI Knowledge and Understanding for aviation safety (HAIKU) project.

Horstmann, G. (2006). Latency and duration of the action interruption in surprise. Cognition & Emotion, 20(2), 242-273. https://doi.org/10.1080/02699930500262878

KNKT. (2014). Komite Nasional Keselamatan Transportasi (2014), Aircraft Accident Investigation Report, Indonesia Air Asia Airbus A320-216; PK-AXC.

Koch, M. (1999). The neurobiology of startle. Progress in Neurobiology, 59(2), 107-128. https://doi.org/10.1016/S0301-0082(98)00098-7

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Monderer, D., & Shapley, L. S. (1996). Potential games. Games and economic behavior, 14(1), 124-143.

Nash Jr, J. F. (1950). The bargaining problem. Econometrica: Journal of the econometric society, 155-162.

Shapley, L. S. (1953). A value for n-person games.

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