Deliverable D6.2

First validation report and demonstrator

Authors: Dong Bach Vo (ENAC)

Major contributors: Alexandre Duchevet (ENAC), Charles-Alban Dormoy (CATIE), Spyrou Evangelos (CERTH), Barry Kirwan (ECTL), Roberto Venditti (DBL), Carl Westin (LiU)

Reviewed by: Vanessa Arrigoni, Simone Pozzi (DBL)

Abstract:

The HAIKU project aims to generate knowledge on intelligent assistants, and to develop AI-enabled 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 results from first validation of WP6 in the HAIKU project – D6.2: First validation report and demonstrator. Based primarily on the European Operational Concept Validation Methodology (E-OCVM), the document provides validation studies descriptions and preliminary project results for each use case.

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

Information Table

Deliverable Number	D6.2
Deliverable Title	First validation report and demonstrator
Version	2.0
Status	Final version
Responsible Partner	ENAC
Contributors	Dong Bach Vo (ENAC) Alexandre Duchevet (ENAC), Jean-Paul Imbert (ENAC), Charles-Alban Dormoy (CATIE), Spyrou Evangelos (CERTH), Barry Kirwan (ECTL), Roberto Venditti (DBL), Carl Westin (LiU), Jekaterina Basjuka (LiU), Karl Johan Klang (LiU), Supathida Boonsong (LiU)
Contractual Date of Delivery	February 29th, 2024
Actual Date of Delivery	February 29th, 2024
Date of Delivery - 2nd version	July 21st, 2024
Dissemination Level	Public

Document History

Version	Date	Status	Author	Description
0.1	11-12-2023	First draft	Dong Bach Vo, Carl Westin	Template for all partners
0.2	09-02-2024	Second draft	All partners	Includes contributions from all UCs
0.3	13-02-2024	Third draft	Dong Bach Vo	Includes contributions from all UC, page layout and styles
0.4	22.02.2024	Draft	Carl Westin, Jekaterina Basjuka, Karl Johan Klang, Supathida Boonsong	Updated UC3 text to final version
0.5	24.02.2024	Draft	Simone Pozzi	Reviewed by DBL
0.6	27.02.2024	Draft	Evangelos Spyrou, Dong Bach Vo	Updated UC6 text to final version.
0.7	27.02.2024	Draft	Alexandre Duchevet	Editorial review by ENAC
0.8	28.02.2024	Final draft	Carl Westin	Editorial review by LiU
1.0	29.02.2024	Final version	Dong Bach Vo, Alexandre Duchevet	Final check by ENAC
1.1	14.06.2024	New amended version	Alexandre Duchevet	Amendment following project officer review
2.0	21.07.2024	Final version	Simone Pozzi	Coordinator's review

List of Acronyms

Acronym	Definition
AI	Artificial Intelligence
ALC	Alicante-Elche Miguel Hernández Airport
AOI	Area of Interest
AP	Autopilot
API	Advanced Program Interface
ASW	Airport Safety Watch
ATCO	Air Traffic Controller
ATHR	Auto-Thrust
ATIS	Automatic Terminal Information Service
CAT	Commercial Air Transport
COMBI	COMmunication in BI-directional way
ConOps	Concept of Operations
CSUQ	Computer System Usability Questionnaire
CTOT	Computed Take-Off Time
DL	Deep Learning
DUC	Digital Assistant for UAM Coordinator
EASA	European Union Aviation Safety Agency
ECAM	Electronic Centralised Aircraft Monitor
EFB	Electronic Flight Bag
EOBT	Estimated Off-Block Time
ETA	Estimated Time of Arrival
FCU	Flight Control Unit
FD	Flight Display
FOCUS	Flight Operational Companion for Unexpected Situations
HAIKU	Human AI Knowledge and Understanding
HAT	Human-AI Teaming
HMI	Human Machine Interaction
IA	Intelligent Assistant
INFOQUAL	Information Quality
INTERQUAL	Interaction Quality
ISA	Intelligent Sequencer Assistant
KPA	Key Performance Areas
LLA	London Luton Airport Company
LTN	London Luton Airport
METAR	Meteorological Aerodrome Report
ML	Machine Learning
MoE	Measures of Effectiveness
MoP	Measures of Performance
ND	Navigation Display
NOTAM	Notice to Airmen
PFD	Primary Flight Display
SPO	Single Pilot Operation
SSD	Single Shot Detector

UAM	Urban Air Mobility
UAS	Unmanned Aircraft System
UATM	Unmanned Air Traffic Management
UC	Use Case
VAL1	First Validation
VAL2	Second Validation

Table of Content

Sommario

1. 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 (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

The goal of WP6 is twofold: to assess whether the project and the IA prototypes are proceeding in the right direction, and to assess the final IA prototypes by providing empirical evidence on their operational benefits. D6.2 First validation report and demonstrator provide the first insights on the design and the evaluation of IAs for aviation safety.

Following up D.6.1, D6.2 is generally based on EUROCONTROL's European Operational Concept Validation Methodology (E-OCVM). However, the six use cases are diverse in terms of domains, needs, end users, implementations, and level of automation. Consequently, the actual evaluation goals, methods, data collection techniques, and metrics vary according to the goals of each use case.

2. Introduction

2.1 Purpose of the document

This document provides the results of the first iteration validation of the six HAIKU use cases (VAL1). The document includes first iteration validation and TRL objectives, description of the user interaction with the AI systems, the evaluation activities, and the results from these activities. In addition, D6.2 includes details about demonstrators (DEM1) for each use case. The demonstrators' videos can be accessed by following the links in Table 1 below.

Table 1. Demonstrator video links

Use case	Video Description	Link to video
#1	FOCUS intelligent assistant demo	https://drive.google.com/file/d/1LwYaI8UFEBF52O1zl86pCFjO0u22Z9K5/
#2	Intelligent Assistant in the cockpit to assist in route planning/replanning	https://youtu.be/i_2zJW7aihs
#3	Digital Assistant for UAM Coordinator (DUC) demo	https://drive.google.com/file/d/1b_vP3zZwlfe3pOdtvHCX1P5WnVd5_AgX/
#4	Intelligent Sequential Assistant (ISA) demo	https://drive.google.com/file/d/1TwrNjVY6h86oFxFo0Mr6EdcU08K7EOqn/
#5	Suite 5 LTN Airport Dashboard demo	https://drive.google.com/file/d/1ZPDjX4m0CakyY0LMCX5aDIOMoh2L7UEC
#5	ENG LTN Airport flight incidents and history browsing demo	https://drive.google.com/file/d/1qCjfwpgfRPYBjBcn7-7av-EgLo9o8_Rp/
#6	Airport virus spreading prevention: passenger tracking system demo	https://drive.google.com/file/d/1VoqkM2slY1nZorrSpbaZ1bUuXHhhH55D
#6	Airport virus spreading prevention: supervision mobile application demo	https://drive.google.com/file/d/1KkzdJvo5Q2HccYCkCsQoPxqNJuF2s_S6
#6	Airport virus spreading prevention: airport air quality interface demo	https://drive.google.com/file/d/1LuAspJYQJckUSMFiviLYK78swPRjb8Sf

2.2 Structure of the document

This document is structured in 9 sections:

- Section 1: Executive summary
- Section 2: Introduction
- Section 3: Use Case #1 – Flight Deck Startle Response
- Section 4: Use Case #2 – Flight Deck Route Planning/Replanning
- Section 5: Use Case #3 – Urban Air Mobility
- Section 6: Use Case #4 – Digital and Remote Tower
- Section 7: Use Case #5 – Airport Safety Watch
- Section 8: Use Case #6 – Airport Spreading Virus Prevention

- Section 9: References

2.3 TRL objectives

The Table 2 below provides an overview of each use case TRL objective and its current TRL.

Table 2. Use cases TRL overview

Function ID	Use Case	Brief Description of Available Digital Assistant or Interface Element	Current TRL	Target TRL
1.1	#1	Startle effect detection	3	4
1.2	#1	Situation awareness augmentation	3	4
1.3	#1	Stress regulation support	3	4
2.1	#2	Communication of high intention for AI reroute planning and alternate choice	3	6
3.1	#3	AI underlying the UAM Intelligent assistant DUC	1	1
3.2	#3	HMI - Interactive touch-based interface and traffic visualisation tool (UTM City) and IA interface.	3	4
3.3	#3	System and convolutional neural networks for image recognition to monitor and process object movement (SOMA-AI) and overlay functionality	3	3
3.4	#3	XAI UI – Storytelling explainer for explaining how intelligent assistant works	1	3
4.1	#4	Service for computing Estimated Time of Arrival and initial trajectory points	3	6
4.2	#4	Optimisation Algorithm to output sequence	4	6
4.3	#4	Explainability for sequence changes	2	6
4.4	#4	HMI - Electronic flight strips and strip board management	3	6
4.5	#4	System's sequence mode (reduced and normal modes)	2	6
5.1	#5	Interactive visualisations to communicate the ML/DL results	3	5
6.1	#6	HMI for Android Application and indoor air quality tool	2	5
6.2	#6	Routing algorithm	2	5
6.3	#6	XAI using SHAP	3	5
6.4	#6	XAI using CLT	1	5
6.5	#6	Wireless Camera and indoor air quality prototypes	3	5
6.6	#6	ML algorithms for classification	3	5

For each UC, more details on the rationale for current and target TRLs are included in the respective sections (section named *UC#n TRL*).

2.4 Validation 1 Progress and Demonstrators

The first iteration of validation activities include a data collection and analysis to assess the IA prototypes with target stakeholders or end users in each use case. All use cases have completed their experimental evaluations. The table below shows when in time, and with how many stakeholders, the different use cases conducted their experiments.

Table 3. Use Case experiment schedule and recruited participants

Use case	Experiment dates	Stakeholder and nr. of participants
#1	28.11.2023 - 06.12.2023	5 Commercial Airline Pilots
#2	15.11.2023 - 26.11.2023	6 Commercial Airline Pilots
#3	29.01.2024 - 02.02.2024	8 Air Traffic Controllers
#4	16.01.2024 - 24.01.2024	8 Air Traffic Controllers
#5	27.02.2024 - 01.03.2024	LLA (3), WIZZAIR (1), NATS (2)
#6	26.01.2024 - 28.02.2024	Injected data simulation, internal validation, small scale staff operation

3. Use Case #1 – Flight Deck Startle Response

3.1 First Validation (VAL1) Plan Summary

In Use Case #1, the main goal is to design and evaluate an IA capable of supporting pilots recovering from the startle and surprise effect. This is done by regulating stress and maintaining or raising the pilots' situation awareness level.

The objectives of the first validation were to collect data to assess the key R&D objectives defined in D6.1.

Table 4. UC#1 key R&D objectives from D6.1.

OBJ-ID	Validation Objective
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-OBJ-02	To assess the acceptability of the solution from the CAT pilots perspective in SPO.
UC1-OBJ-03	To assess the feasibility and integration of the solution in a relevant operational environment.
UC1-OBJ-04	To assess the effectiveness and efficiency of the assistant support in a relevant operational environment.
UC1-OBJ-05	To assess the generalisation of the solution to multiple different scenarios.

The design of the IA startle recovery features was guided by the performance targets defined in D6.1, which are reminded in Table 5 below.

Table 5. UC#1 performance targets

KPA	Category	KPI
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.

The European Union Aviation Safety Agency (EASA) provides guidance on Human-AI Teaming (HAT) with six categories describing Human-AI partnerships. UC1 Intelligent Assistant, named FOCUS, aims at providing support for the following categories:

- 1B – Cognitive assistant, equivalent to decision support system

- 2A – Cooperative agent, able to complete tasks as demanded by the operator

3.2 Deviation from Deliverable 6.1

In the deliverable D6.1, three exercise scenarios were proposed: lightning strike on final approach, shifting cargo at take-off, overspeed upon windshear occurrence in climb. Due to ENAC's A320 simulator technical limitations the scenario 3 (windshear) could not be implemented.

3.3 Assistant Features

Three support functions are currently integrated to the IA in UC1: a startle and surprise detection module; a stress regulation support feature; and a situation awareness support feature. In addition, an assistant monitor and a control panel is provided on the pilot's electronic flight bag (EFB). The complete architecture and design rationale of FOCUS are provided in the deliverable D4.4.

Startle and surprise detection module

The role of the startle and surprise detection module is to monitor the pilot's physiological data and eye gaze behaviour to identify any situation of startle and/or surprise. Upon startle detection, the assistant notifies the pilot that it will start recovery procedures by displaying an image on the primary flight display (PFD) and ECAM for 500 milliseconds. Then it performs stress regulation support first, and situation awareness support next. In the case where only surprise is detected, the assistant performs situation awareness support only. The auto-recovery can be cancelled by the pilots within 5 seconds after startle detection.

Currently, the system is not yet capable to detect startle and surprise states accurately. The detection of such events was simulated for VAL1. However, the startle and surprise detection module is being developed and refined in parallel. The physiological and behavioural data collected during VAL1 will support the improvement of the models for detecting startle and surprise in the cockpit.

Stress regulation support

The assistant can limit the negative effects of acute stress experienced by the pilot following two approaches.

The effect of deep breathing on stress has been extensively demonstrated (Perciavalle et al., 2017). Breathing can be slowed down to a constant rate to increase relaxation and reduce stress levels (Schlatter et al., 2022). However, reaching an optimal breathing pace requires self-awareness and control. Therefore, the assistant will help pilots to reach cardiac coherence, a state where the heart rate variability is better controlled, to reduce the effect of stress. To guide pilot's deep breathing, a green halo visual effect on the PFD and a green ambient light in the cockpit are used (Figure 1). While as the light intensity on the PFD and in the cockpit increases, the pilot breathes in, as the light intensity decreases, the pilot breathes out. This procedure is based on the Heart-Focus Breathing technique that is a common step to increase cardiac coherence where each breathing cycle lasts for 10 seconds (McCraty & Zayas, 2014).

Research has shown that constant low heart rate feedback can significantly reduce anxiety (Sun et al., 2023). Experiencing tactile feedback with a simulated heartbeat at 60 beats per minute can reduce perceived anxiety (Costa et al., 2016) and heart rate after some physical efforts (Choi & Ishii, 2020). The second approach to limit the effect of stress is to provide pilots with vibrotactile feedback. The pilot is equipped with a device on the wrist, which provides a simulated heartbeat through vibration every second, i.e. 60 beats per minute, when the stress regulation support is active. It is worth noting that any wrist-worn vibrating device such as common smart watches could be used for this type of stress regulation intervention.

Figure 1. Deep breathing guiding light in the cockpit

Situation awareness support

The situation awareness support ensures pilots update their awareness of the aircraft status. When the support is active, the assistant highlights the instruments on the PFD that need to be checked. When highlighted, the instrument is wrapped with a coloured rectangular box. As the pilot glances at the instrument, the highlight disappears. The assistant uses pilots' gaze tracking data and flight parameters to assess the urgency of the flight parameter update. Two levels of urgency are communicated: amber as a warning if the instrument has not been checked recently, red as an alert if the instrument needs to be checked immediately (Figure 2).

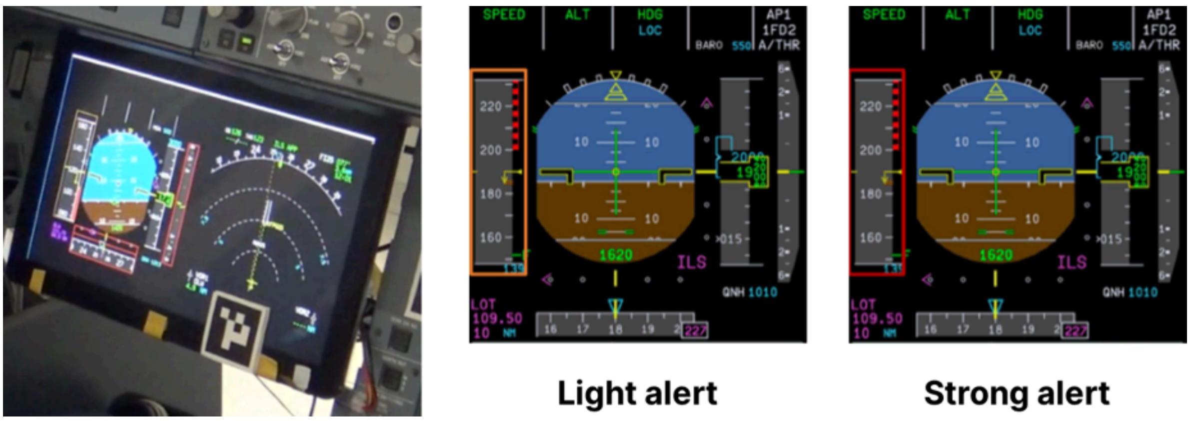


Figure 2. Situation awareness support and alert levels.

These levels are represented by two distinct designs on the PFD, the navigation display (ND), and the electronic centralised aircraft monitor (ECAM) (Figure 3). In order to estimate the pilot's situation awareness (SA), eye tracking data are used to analyse its visual path and a situation awareness score is calculated. The SA score for each aircraft parameter is degraded as a function of time since last viewed by the pilot, it will also be changed if the parameter changed significantly according to the flight phase since last viewed. For each parameter, according to the flight phase, a min/max threshold has been set following interviews with commercial pilots. The eye tracking system helps FOCUS to know whenever a parameter has been seen. More details on the SA score computation can be found in the deliverable D4.4.

Figure 3. Locations of situation awareness support displays

Assistant monitor and control

The assistant support functions can be controlled using an application from the pilots' EFB placed on the window side of the piloting position. Each support function can be activated and disabled manually on demand. This always provides full control of the assistant to the pilot. The monitor also displays information which helps the pilots always understand the operation of the intelligent assistant.

Firstly, the pilot can monitor his heart rate through the assistant. This piece of information is useful to raise one's self-awareness of his own physiological state and to create an opportunity to enable stress regulation support if needed. It also allows the assistant to provide information about its functioning and some explanation for self-activation to the pilot.

Secondly, the pilot can monitor the relevancy of his situation awareness. A global situation score shown on the display indicates whether flight parameters reading is required. In addition, a widget mimicking the PFD suggests which areas of interest should be checked to increase the global situation awareness score. These contribute to explain the assistant functioning and identify directed opportunities to improve the pilot's situation awareness.

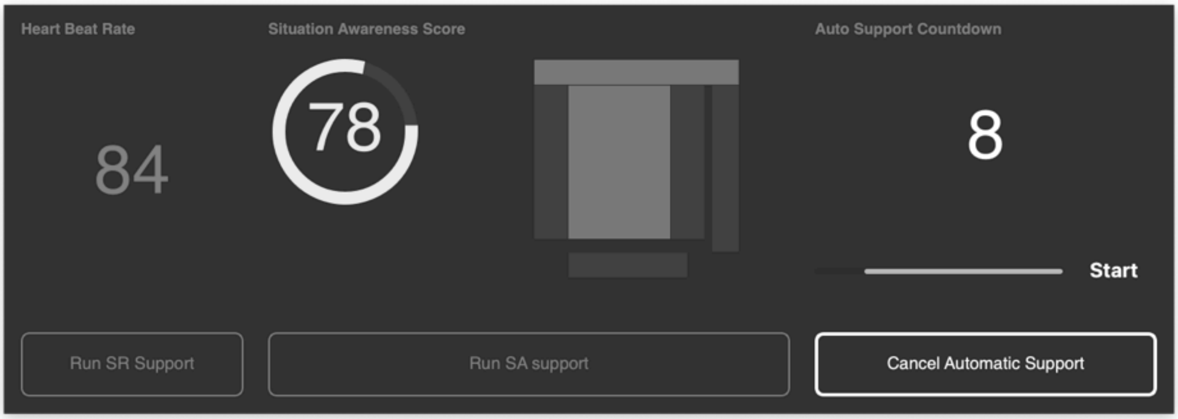


Figure 4. Assistant monitor: stress regulation, situation awareness and auto-support and control.

Finally, the monitor provides pilots with the ability to cancel the automatic support execution when the intelligent assistant has identified a situation of startle. This ensures that when support is not required, the support system will not interfere with the piloting tasks.

3.4 UC#1 TRL

In UC1, the FOCUS assistant is composed of 3 main functions recalled below with their current TRL:

- Startle effect detection, TRL3
- Situation Awareness augmentation, TRL3
- Stress regulation support, TRL3

The startle effect detection module performances were evaluated in laboratory settings, showing encouraging results. However, the performances were not good enough for using the module in the simulator for Validation 1 activities. The startle effect detection function was not implemented for Validation 1. The module demonstratively works in laboratory settings and is an advanced proof of concept, it therefore validates its TRL3. With the data collected during validation 1, we hope to achieve a good enough startle effect detection to validate TRL4 for Validation 2.

The situation awareness function and the stress regulation support were implemented in the simulator for Validation 1. They achieved TRL3 as it is working well in laboratory. Validation 1 and Validation 2 will permit to reach TRL4 as these functions will be tested and evaluated with end-users in a relevant environment namely the ENAC A320 simulator.

3.5 Scenarios description

Due to ENAC's A320 simulator technical limitations, Scenario 3 described in D6.1 could not be implemented.

Scenario 1: Lightning strike during 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 of the aircraft lead to automations disconnection.

The scenario begins during the holding pattern for Runway 25 at Orly airport. An event triggered automatically by the platform software manager will provoke a low fuel warning in the cockpit (Figure 5).

The pilot is supposed to say out loud 'Master caution, Low fuel', remove the auditory warning by pushing the Master caution button on the flight control unit, and ask the air traffic control to integrate the final approach immediately by saying "ENAC2023 request radar vectoring for immediate approach". In addition, the pilot should also report a low-fuel emergency situation by declaring "PAN PAN" or "MAYDAY MAYDAY" on the radio. Following this message, the air traffic controller will give a heading to ENAC2023 in order to integrate the Runway 25 approach. ENAC2023 is then supposed to call back when the Localizer is captured. This low fuel situation, while increasing the urgency of the situation and raising the stress level of pilots, is a necessary "scenario trick" to force participants to land.

At 12 Nm from the runway (t0), the lightning strike and thunder are triggered in the cockpit. Following this event, autopilot (AP), auto-thrust (ATHR) and flight director (FD) are disconnected. Since the startle detection is not integrated to our platform for VAL1, the startle detected event is triggered automatically by the platform software manager, which will lead the pilot into a startled state.

The startle detected event starts the emotion regulation support function after 5 seconds. This function lasts for 30 seconds. 15 seconds after the startle event, the situation awareness function starts and lasts for at least 90 seconds, depending on if the pilot's situation awareness is adequate. The pilot can start or stop both functions manually at any time on the assistant user interface. During the first 5 seconds after startle detection, the pilot can stop the assistant before its activation.

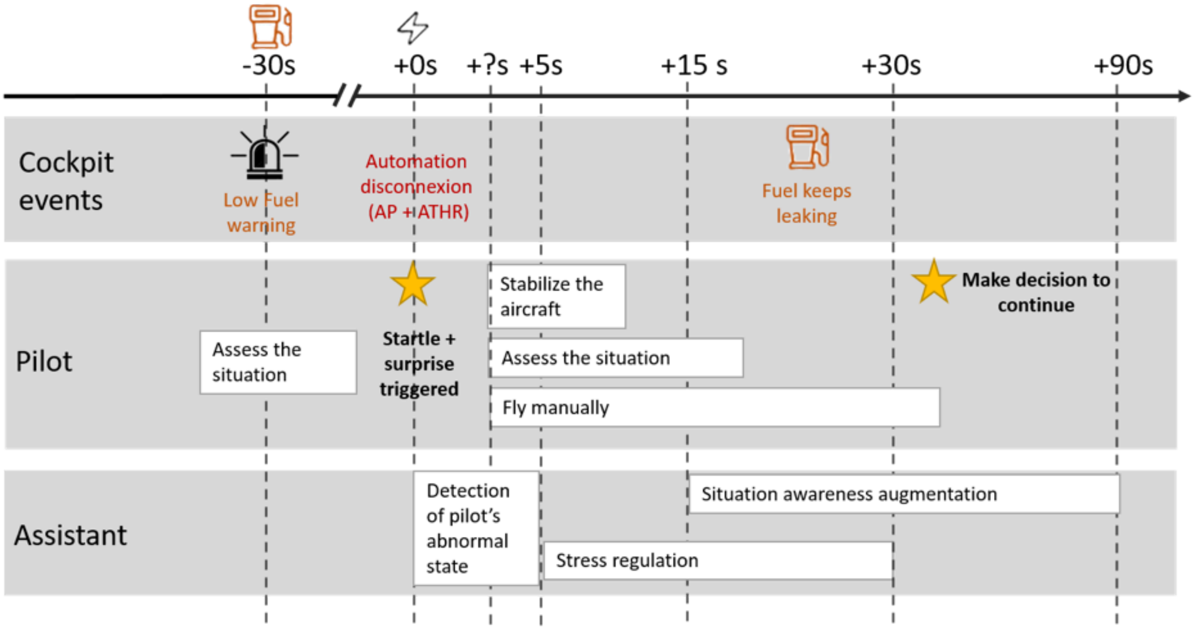


Figure 5. Lightning strike scenario script

Scenario 2: cargo shift during take off

At the beginning of this scenario, the aircraft is aligned on LFBO Runway 32R. Weather is clear with no wind. Shortly after take-off (when the aircraft reaches a pitch of 7.5°), a 6 tons cargo gets loose and moves from the front to the rear part of the cargo bay resulting in a shift of centre of gravity (Figure 6).

As a result, a strong pitch up moment is observed, up to a pitch of 20°, that will supposedly surprise the pilot. Because of the strong pitch, the AP and FD functions cannot be engaged by the pilots as usual, and the pilot is forced to react quickly and manually control the aircraft.

As a consequence, the aircraft manoeuvrability is impacted, and the pilot is supposed to ask for an adapted trajectory to come back and land on the airport. 5 seconds after the detection of the surprise event, the situation awareness support starts and lasts for 90 seconds if the pilot's situation awareness is correct.

The pilot can start or stop the SA support manually at any time on the assistant user interface. During the first 5 seconds after the startle occurrence detection, the pilot can stop the SA support before its activation.

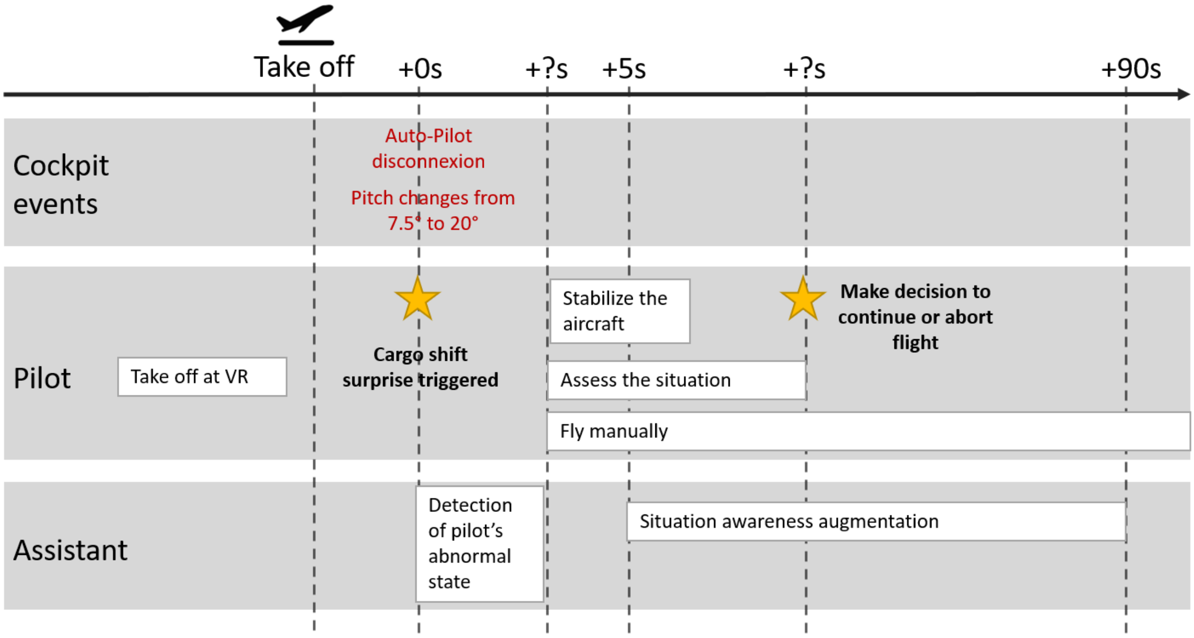


Figure 6. Cargo shift scenario script

3.6 UC#1 Validation 1 Procedure

Upon consent form completion, each participant was first introduced to the research objectives. The experimenters then presented the goal and the proceedings of the study, and specifically, the context, the AI features and the two flight scenarios of the study without disclosing the startling and surprising events (lightning strike and cargo shift). In each scenario, the flight was conducted in SPO. Each flying session took place in an Airbus A320 simulator built at ENAC.

Next, the participant was equipped with the physiological sensors and was invited to sit in the A320 simulator. Upon eye tracking calibration, one of the experimenters performed a walkthrough of the IA functions and invited the participant to experience the stress regulation and the situation awareness support. The participant could ask any question about the IA at this stage.

When ready, the participant began the training phase. The training phase started with exploring the cockpit surroundings for at least 20 minutes to give enough time for the participant to familiarise with the simulator, the controls' location and sensitivity, and the simulation view rendering. In addition, the participant was allowed and encouraged to request support and experience the assistant functions during the training. When the participant felt comfortable enough to proceed, a baseline scenario was started which consisted in taking off, performing a short flight and landing back to the airport.

Upon completion of the training, the participant performed the two validation scenarios. During each scenario, photoplethysmograms, electrocardiographs, electromyographs, galvanic skin responses (GSR), respiration rates, gaze, participants' view and participants' face were recorded. At the end of the scenarios, the participant was debriefed about the experience and performance during the flight. Between each scenario, the participant filled a Likert-scale questionnaire to assess the subjective perception on the performance, the usefulness and the understanding of the IA.

Finally, when the two scenarios were completed, the participant was invited to debrief about the stress regulation and situation awareness support functions first. Then, the participant was questioned about the IA through a semi-structured interview in which usability, improvements, AI initiative and trust was discussed.

Five experienced male pilots were recruited for the validation study (mean age = 44.8 SD = 7.5, mean number of flight hours=4600, SD = 5672). The gender was not balanced, considering that female pilots represent 5.8% of the pilot's overall population. All performed the lightning strike scenario. Three performed the cargo-shift scenario in addition.

3.7 UC#1 Validation 1 Results

Although it is not possible to report statistics analysis given the sample size of the study, the questionnaires result analysis can provide some subjective perception tendency among the participants.

The results from the first validation show that:

- the designed scenarios for evaluating the UC#1 intelligent assistant triggers startle and surprise
- the participants were overall positive with using the assistant
- the participants are aware when FOCUS is active
- the ambient light in the cockpit and on displays is seen as a stress marker
- the physiological data monitoring should not be displayed on the EFB
- tactile heartbeat stimulation is a promising solution to regulate stress for pilots
- FOCUS is found useful especially in manual piloting
- FOCUS provides too many visual alerts on the cockpit displays.

In the next sub-sections, we report the results of the First Validation study in more detail.

General feedback

First of all, **the scenarios successfully triggered startle and surprise in the participants**. On a scale from 1 (not startled or surprised) to 10 (very startled or surprised), participants reported an average score of 7.0 for startle and an average score of 7.6 for surprise in the lightning strike scenario. The cargo shift scenario was deemed less intense with a startle average score of 3.7 and a surprise average score of 5. Physiological data and facial expressions confirmed that all the participants were actually startled or surprised during the lightning strike scenario (Figure 7). Signs of stress after the surprising event were observed in physiological data in the cargo shift scenario for all the participants. For example, P4 commented: "I was sort of surprised and trying to figure out what is working and where I am, what direction am I going?".

The assistant was generally well received by the participants. P1 stated: "That is a good approach, that is a good idea". The participants thought that the assistant allowed them to maintain a good situation awareness and the awareness guidance was overall relevant. P2 also thought that the assistant was not restrictive. The system actions and purpose were well understood by the participants, and the assistant was thought easy to interact with. Participants felt somewhat confident to work with the assistant. It is worth noting that P1 said that he would trust the system. That being said, the participants felt unsure about the benefits of the assistant to limit the detrimental effect of startle and surprise, and its usefulness when unexpected events occurred (i.e. in situations of surprise).

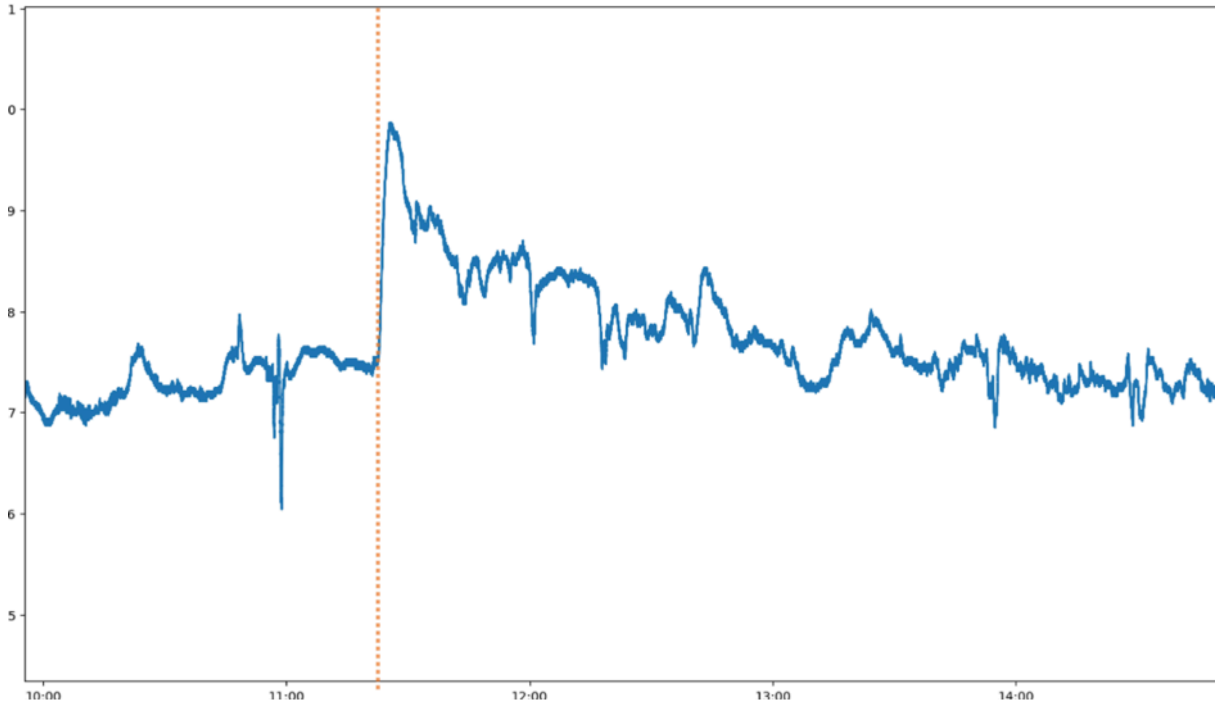


Figure 7. Galvanic skin response of a participant following the lightning strike event

Assistant start

Participants were aware when the assistant had been triggered off. P1 stated: "I saw the small icon" and P4 reported: "I saw the two persons symbol coming up". The visual alert was seen as a cue for abnormal situations. It disturbed P4 by disrupting his visual scan but allowed him to "take a step back" to what was happening. He added: "I am thinking that might have been helpful because it kind of told me that, now, something is wrong". P5 underlined that this feature was reminding him of other features that he already knows such as the traffic collision avoidance system (TCAS) blue notification[1], which is displayed on PFDs when a TCAS Resolution Advisory maneuverer may be necessary.

Stress regulation support

[1] Safe Handling of TCAS Alerts, AIRBUS. https://safetyfirst.airbus.com/safe-handling-of-tcas-alerts/

The breathing guidance lights were seen as an indicator of stress level during the exercise (P1). By seeing the light, P4 was reminded to breathe deeply. However, participants highlighted the lack of availability to perform the breathing procedure. For instance, P2 stated: "You need availability to focus on your cardiac rhythm" and P3 added: "I was thinking about it during the last [led] cycle, then I tried to adapt my breathing. Before that, I don't think I was not aware of my breathing nor adapting my breathing". Although P2 saw the lights, he thought that focusing on his breathing while dealing with the aircraft automation failures was quite difficult.

The tactile heartbeat simulation was not found uncomfortable. As a matter of fact, some participants did not even notice the tactile feedback during the exercise even though they experienced it during the training phase. Pilots were eager to know whether the tactile feedback had a real impact on their heartbeat. P2 thought that pilots could benefit from tactile feedback as it was "transparent" (i.e. unintrusive) to him and to other pilots, and that it did not have any impact on his workload. He added that the technology may be promising if it can be compatible with existing pilot smartwatches.

Finally, it is worth noting that **none of the participants found the physiological monitoring display useful**. Even though P3 was monitoring his physiological status at the beginning of the exercise, he stopped when the assistant support started. The other pilots reported that they did not look at the electronic flight bag during the exercises. They reported indeed that they did not have the cognitive resources to check the physiological monitoring display once the emergency was declared. They had to focus on handling the emergency and control the aircraft.

Situation awareness support

The situation awareness support was useful to participants, especially when the autopilot failed in our exercise. The assistant managed to draw pilots' attention towards specific parameters. P5 commented: "I missed the speed change. I was glad that the assistant told me to look at the speed". P4 thought the benefit of the situation awareness support was to "increase the sampling rate" to acquire flight information. He thought that it may have focused too long on the navigation display or on some other information. P2 thought it helped to recentre the visual path. P3 said that the red boxes helped him to check the right pieces of information, even though this is what he was already planning. He reported: "I think I would have done it. But it allowed me to save time". P1 also commented on the potential for such assistance to support pilot flying aircraft that they are not familiar with. With more training with the assistant, the participants thought they could follow its guidance better.

However, **they also warned about the potential distraction that could result from the PFD red highlighting boxes**. Because the pilots did not look at all the instruments, more boxes started to appear on the PFD. They thought that this was overwhelming and confused them about what instrument to look at first. P5 reported: "at the end there were too many zones to look at" and criticised the use of the red colour: "I have noticed that when amber changes to red, it creates a strong distraction".

The pilot's inability to make some boxes disappear upon glancing at it was partly due to the system's lack of performance in recognizing Area of Interest (AOI) in the cockpit. Particularly, it failed several times to detect the AOIs associated with the heading and the localizer deviation on the PFD resulting in constant red boxes around these zones even though pilots were staring at them.

The reliability of the assistant was questioned. In the thunderstorm exercise, autopilot 1 and 2 are expected to fail after the thunderstrike hit the aircraft. Because the pilot had been given erroneous information because of the automation failure in the scenario, P2 was not sure whether he could still rely on the assistant guidance. He was also worried that the assistant would focus his attention on an erroneous piece of data.

3.8 Improvements and future work

Since the goal of VAL1 was to collect feedback on the use of our assistant, we did not evaluate the action of cancelling the auto-activation if not needed by the participant. For VAL1, the delay before the assistant executes support was fixed to 5 seconds. Based on our observations during the exercises, the delay duration was too short. However, **it is not clear how long this delay should last**. Not only it should allow the pilot to cancel the auto-support before it launches but also provide support as soon as possible to pilots, as startle effects consequences are at the strongest immediately after the stimulus. This may be investigated further.

Although the breathing guidance lights were not invasive to participants during the exercises, **the demand of resolving aircraft automation failures did not give any cognitive space to pilots for performing deep breathing**. None of the participants froze after the thunder strike stimulus indicating that none of them were incapacitated during our study. This may have contributed to ignoring the breathing support. Deep breathing is key to reach cardiac coherence and reduce stress, therefore we will continue to investigate ways to increase pilot's attention to breathing and explore new forms of breathing support to help pilots reduce their stress.

Along the same lines, **the visual biofeedback cues displayed on the electronic flight bag were not relevant to the pilot's tasks**. Since understanding failures and building situational awareness requires the pilots' attention to the flight and navigation displays, none of the pilots looked at the electronic flight bag on the side panel after the stimulus had occurred. **The breathing light support was seen as a stress indicator**. Previous research has shown that ambient light can convey information about stress levels and facilitate the control of heart rate response to stressful situations (Yu et al., 2018). Ambient lighting in the cockpit could facilitate stress regulation without requiring the pilot's direct visual attention. This will be investigated further.

The simulated tactile heartbeat feedback was well received by the participants. They highlighted several benefits such as being inconspicuous and undemanding. The participants were interested in verifying the impact on their stress of such an intervention. Even though research has found an effect of low heartbeat tactile stimulation on perceived stress and heart rate, we could not provide the participants with a sound answer as we did not investigate this research question in this study. Given the potential benefits of this approach, it seems that it may be worth investigating the effect of such interventions in the cockpit.

Although the situation awareness support did help participants to pilot the aircraft manually by highlighting the relevant pieces of information, the system can be improved. As performance issues of the assistant resulted in the permanent highlight of 2 of the 11 areas of interest, some participants felt a bit overwhelmed. It underlined the fact that **too many eye-catching alerts on cockpit screens is not an acceptable design and that a prioritisation of information highlighting could be appropriate** as P4 suggested. Another significant drawback was the distracting nature of the strong red alerts on the pilot flight display. Participants suggested using auditory messages such as "speed!" to bring attention to the speed, as they are already used to such callouts done in today's operations by the PM. In addition, oral alerts may provide a sense of pilot monitoring presence in a single pilot setting. We will investigate both priority queues and oral alerts further in the next iteration.

Finally, it is worth mentioning that pilots also shared some ideas for the future of the IA in the cockpit. For example, P1 envisioned an assistant that could adapt to different piloting profiles. He also mentioned a voice assistant that could understand pilots' vocal instructions. Identifying and analysing system failures in single pilot operations may be more difficult, therefore P2 hoped that the assistant would support him in this task to relieve pilot's stress even more.

3.9 UC#1 Validation 1 Conclusion

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

Table 6. UC#1 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 the participants but the performance of such	Because of being inconspicuous and undemanding and easily deployable in the cockpit through existing technology, this may be one of the	Evaluate the impact of simulated tactile heartbeat feedback on stress regulation.

an approach was not investigated in VAL1.	best solutions for regulating stress in the cockpit	
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.

4. Use Case #2 – Flight Deck Route

4.1 Introduction/Background – Use Case #2

In the aviation domain, pilots face a myriad of challenges during flights, ranging from adverse weather conditions to technical failures and dense air traffic. Limited cognitive resources in the cockpit can sometimes hinder the precise assessment of flight optimization and risk management, especially when multiple factors converge.

UC2 is based on a cockpit Intelligent Assistant (IA) to help flight crew re-route an aircraft to a new airport destination due to deteriorating weather, taking into account a large number of factors (e.g. remaining fuel available and distance to airport; in-route turbulences, connections possible for passenger given their ultimate destinations; etc.). The flight crew remains in charge but communicates/negotiates with the IA to derive the optimal solution.

UC2 explores the implications of integrating artificial intelligence (AI) and machine learning (ML) in aviation, particularly to alleviate pilots' cognitive burdens and foster collaborative human-machine interactions.

The purpose here is to provide detailed insight into the methodology used and the results obtained during the initial validation of the UC2 project. By doing so, it sheds light on the possibilities offered by AI to address operational challenges faced by pilots, particularly in an increasingly complex aviation environment.

For reference, the performance targets table from the Deliverable 6.1 is reported below in Table 7. More details of our strategy and approach are given in Deliverable 6.1.

Table 7. UC#2 performance targets

KPA	Category	KPI
Mission Safety	MoE	Safety margin index [ ]. (prescribed activity) (ex.: fuel at destination)
Mission Commercial Performance	MoE	Operational impact index [function of cost efficiency, punctuality, passenger experience, …, TBD with SME]
Decision Quality	MoP	Decision quality index (function of aspects taken into account, TBD with SMEs) Implementation Feasibility index (function of time to decide, time to implement, ATM considerations…, TBD with SMEs)

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

Regulatory Acceptance	MoE	Beyond 2030: Self evaluation against the SoA, and an acceptance of SMEs from regulatory organisations.
Social Acceptance	MoE	Usage of the assistance due to perceived usefulness (reliability, trust, performance…)

Note: In this validation, we made some preliminary hypotheses about the technical requirements and Human-IA interactions. These hypotheses are based on recommendations provided by The European Union Aviation Safety Agency (EASA) about AI and its potential impacts upon aviation operations and practices. EASA's recent guidance on Human-AI Teaming (HAT) suggests six categories of future Human-AI partnerships. As it is presented in section 4.2.3, we made a focus on the categories in which the human work together with the IA:

- 1B – Cognitive assistant (equivalent to decision support system)
- 2A – Cooperative agent, able to complete tasks as demanded by the operator
- 2B – Collaborative agent – an autonomous agent that works with human colleagues, but which can take initiative and execute tasks, as well as being capable of negotiating with its human counterparts

Research question:

UC2's proposal is to work on mission management using high-level Operational "Intentions" (e.g. green operations, quality of service, punctuality, fuel and global cost reduction, etc.). Intentions involve mental activities such as planning and forethought, they can be declared and clearly defined, while in other instances can be undeclared or masked, making them sometimes complex to identify.

To answer this question, we propose to focus on these 3 objectives in this initial validation:

UC2-OBJ-02: What are the key features for each type of assistance (decision support - 1B, cooperative - 2A, collaborative - 2B) that enable teamwork requirements assurance and effectiveness?

UC2-OBJ-04 2A variant: HAT cooperative teaming improves decision making process for on air re-route situation vs. decision support assistance

UC2-OBJ-05 2A variant: 2B variant: HAT collaborative teaming improves decision making process for on air re-route situation vs. HAT cooperative teaming

In COMBI (COMmunication in BI-directional way) the pilot uses "intentions" to communicate with the IA. Intention is a mental activity oriented to achieve the goal in a specific way. Once you have an intention in mind—let us say, for instance, when we want to use our GPS navigator, we have the objective to go to the airport. There are different ways to go to the airport. The choice will depend on our intentions. In this context, several intentions could be proposed: being faster, being cost efficient, being green, etc. Thus, our GPS will be able to provide us with some solutions based on these intentions.

4.2 UC#2 Material And Methods

Experiments

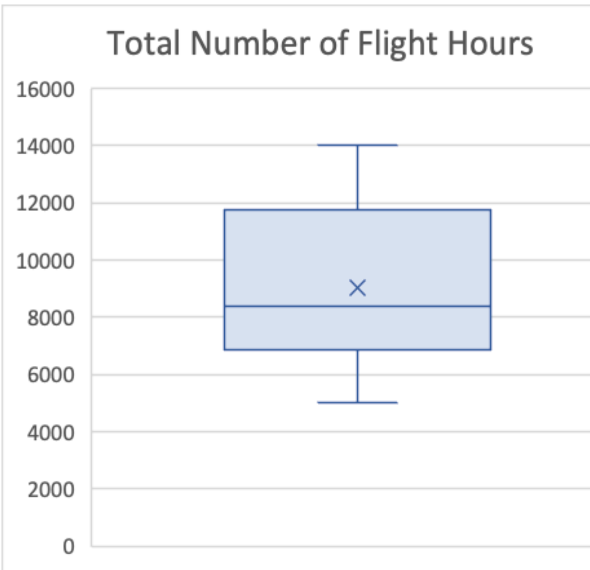
Six airline pilots were engaged in testing three levels of pilot assistance mock-up based on AI classification from EASA: 1B, 2A, and 2B. The experiment using an interactive PowerPoint, simulated a regional flight in Europe, specifically from Marseille to Munich.

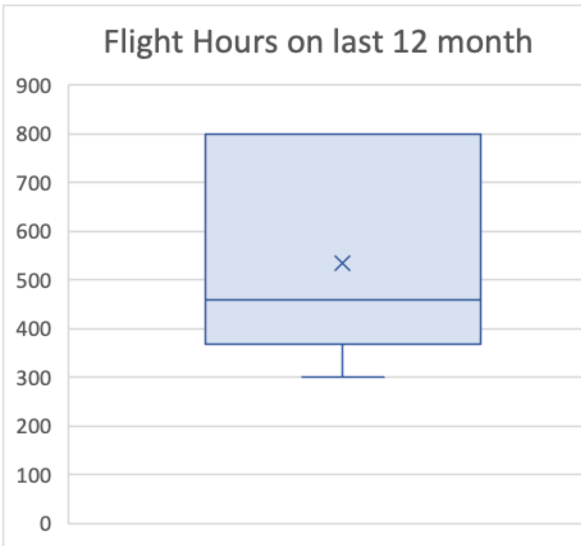
In order to determine the need for an IA according to the complexity of the situation, we have created two types of scenarios with lower and higher complexity. During the flight scenario, the participants encountered the need for in-flight rerouting to avoid adverse weather conditions (less complex scenario), and as second step the closure of the destination (i.e. Munich) airport due to inclement weather, necessitating the selection of an alternate airport (more complex scenario).

Participants

Pilot characterization (demographic questionnaire)

The panel comprised six male commercial airline pilots selected from diverse aviation backgrounds. Their ages ranged from 39 to 59 years, with one participant aged between 39 and 40, four participants aged between 40 and 49, and one participant aged between 50 and 59. On average, the pilots accumulated 9,045 flight hours throughout their careers (SD = 3,099), with an average of 535 flight hours within the last 12 months (SD = 215), reflecting variability in flight experience and recent activity among the participants. One female airline pilot participated in the elaboration of the use case scenario.





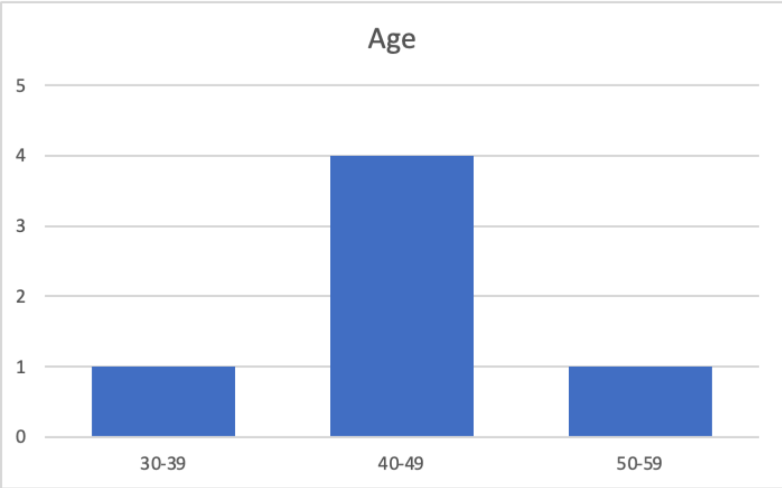


Figure 8. Total flight hours, flight hours last 12 months and pilots' age

The pilots were tasked with evaluating three types of AI assistants and their associated interfaces using the "Cognitive Walkthrough" method. This method, rooted in human factors engineering, aims to systematically evaluate the usability of concept by simulating user interactions and decision-making processes.

In the Cognitive Walkthrough, pilots were presented with simulated flight scenarios and guided through tasks involving the AI assistants and interfaces. As they navigated through each interface, pilots were prompted to verbalise their thought processes, interactions, and decision points. Researchers observed and documented pilot behaviours, identifying potential usability issues, cognitive bottlenecks, and interface complexities.

By leveraging the expertise and insights of experienced pilots, the Cognitive Walkthrough method provides valuable feedback for refining AI assistants and interfaces in aviation contexts. This structured approach facilitates the identification and mitigation of usability challenges, ultimately enhancing the effectiveness and user experience of AI technologies in aviation operations.

Test modalities

The three modalities tested are as follows:

- Support to decision (1B): 3 routes are proposed linked to pilot's intentions, and the associated flight plan (FP) can be implemented in FMS on request. The 3 routes maximise the first intention.
- Cooperative Assistant (2A): low-impact threats - 1 route linked to pilot's intentions, implemented on request /other cases - work same as 1B. The route maximises the first intention.
- Collaborative assistant (2B): 2 routes are proposed - 1 called "Least negative impact", and 1 called "Best Compromise". The respective FP can be implemented in the FMS on request.

- "Least negative impact": the other intentions may lose value, but the loss will be shared as evenly as possible between the two intentions.
- "Best compromise": The loss of value for the other two intentions will be assessed against the gain on the first intention. We will accept low losses for low gains.

In all solutions proposed by the assistants, safety is always a priority and guaranteed. For the reroute case, the new route is always safe, ensuring that the aircraft arrives at destination with a functional machine, safe crew and minimum legal fuel. For the choice of the alternate airport, the landing with safe performance, safe machine, crew, minimum legal fuel etc are also always ensured. This information is displayed on the HMI as a grey bar to show that it is not changeable and set at 100% (see section 4.2.8).

Trials calendar

To randomise the order of assistance, a permutation was designed as shown in the table below. Two pilots were on each serial. The serial 1 started with assistant 1B in a Winter scenario, then 2B in a Summer scenario, and 2A in an Autumn scenario. The serial 2 started with the 2A assistant and Serial 3 with the 2B assistant and finished with 1B.

	CW1		CW2		CW3	
Scenario/Ass. lvl	Scenario	Ass. lvl	Scenario	Ass. lvl	Scenario	Ass. lvl
	Rerouting	Diversion	Rerouting	Diversion	Rerouting	Diversion
Serial 1	W	1B	S	2B	A	2A
Serial 2	S	2A	A	1B	W	2B
Serial 3	A	2B	W	2A	S	1B

Table 8. Permutation order for the 3 serials, with season (Summer, Autumn, Winter) and assistant level (1B,2A,2B) (CW=Cognitive Walkthrough)

Experiment schedule

The first part of the experiment started with a presentation of the HAIKU project, the objectives and the whole experiment process (see Figure 9). We ensured that the consent form was signed before starting the recording. Then the participants were invited to answer the questionnaire about trust in AI.

A presentation of the assistant COMBI was then realised, with the explanation of communicating with the assistant via the three high level intentions: Pilot cognitive comfort, Passengers comfort and Profitability (description of these operational intentions will be provided in deliverable D4.5). A presentation of the interface and the interactions were then demonstrated and tested by the participant (see Figure 10) on the interactive PowerPoint. The pilot can select one of the six triangles to communicate to the assistant the intention chosen.

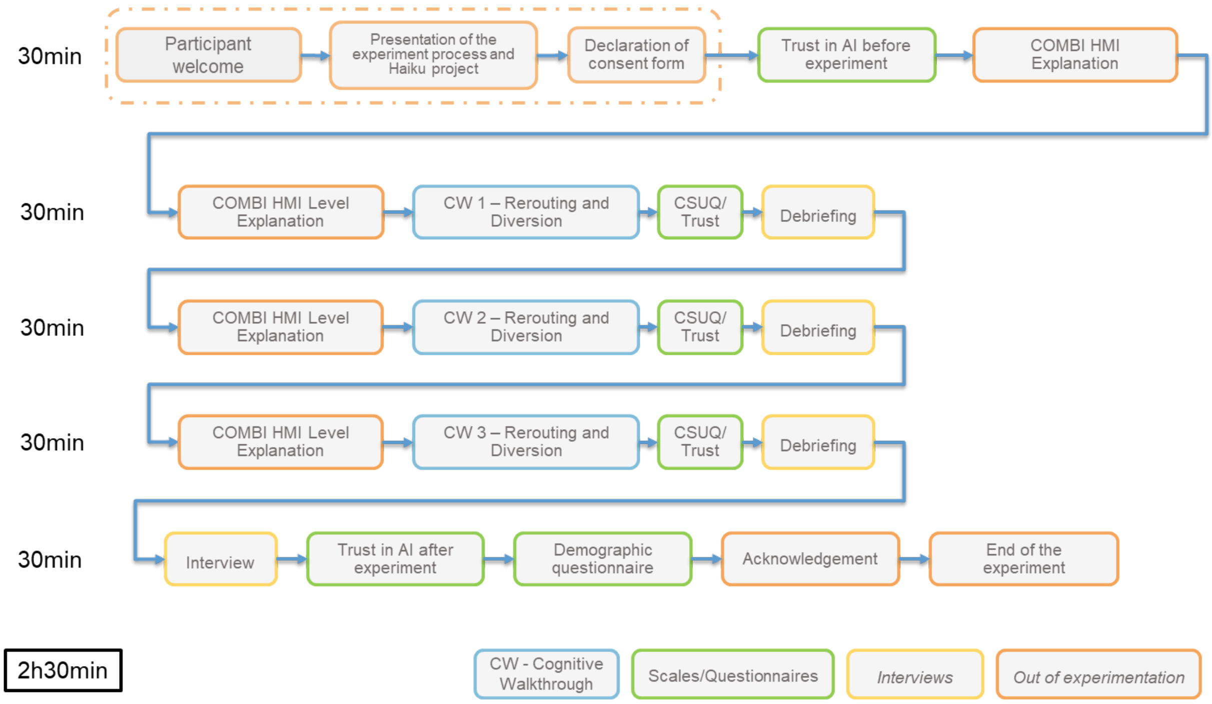


Figure 9. Experiment chronology

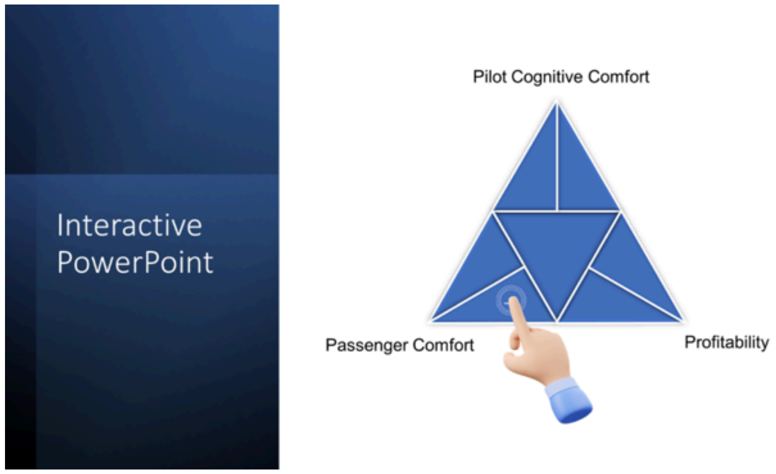


Figure 10. Interactive PowerPoint as mock-up, communication of the pilot's intention to the assistant.

For example, in Figure 11, the intention selected is displayed in green. The triangle on the left prioritises Pilot Cognitive Comfort, then Profitability (same side of the triangle) and at last the passenger comfort. As safety is always ensured by the assistant, a green bar is displayed next to the triangle and also under the intention's figures in grey (see Figure 14).

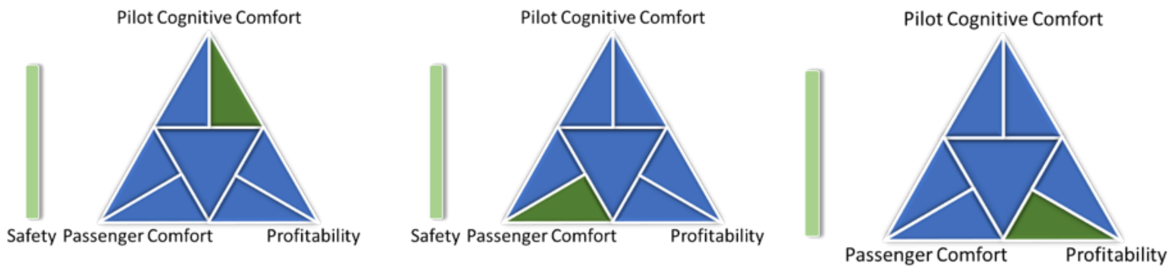


Figure 11. Selection of one intention: left Pilot Cognitive comfort, in the middle Passenger comfort, and at the right the intention selected is profitability.

The format of PowerPoint was chosen to be able to do the remote interviews with the pilots and be able to easily share the interface and make it interactive.

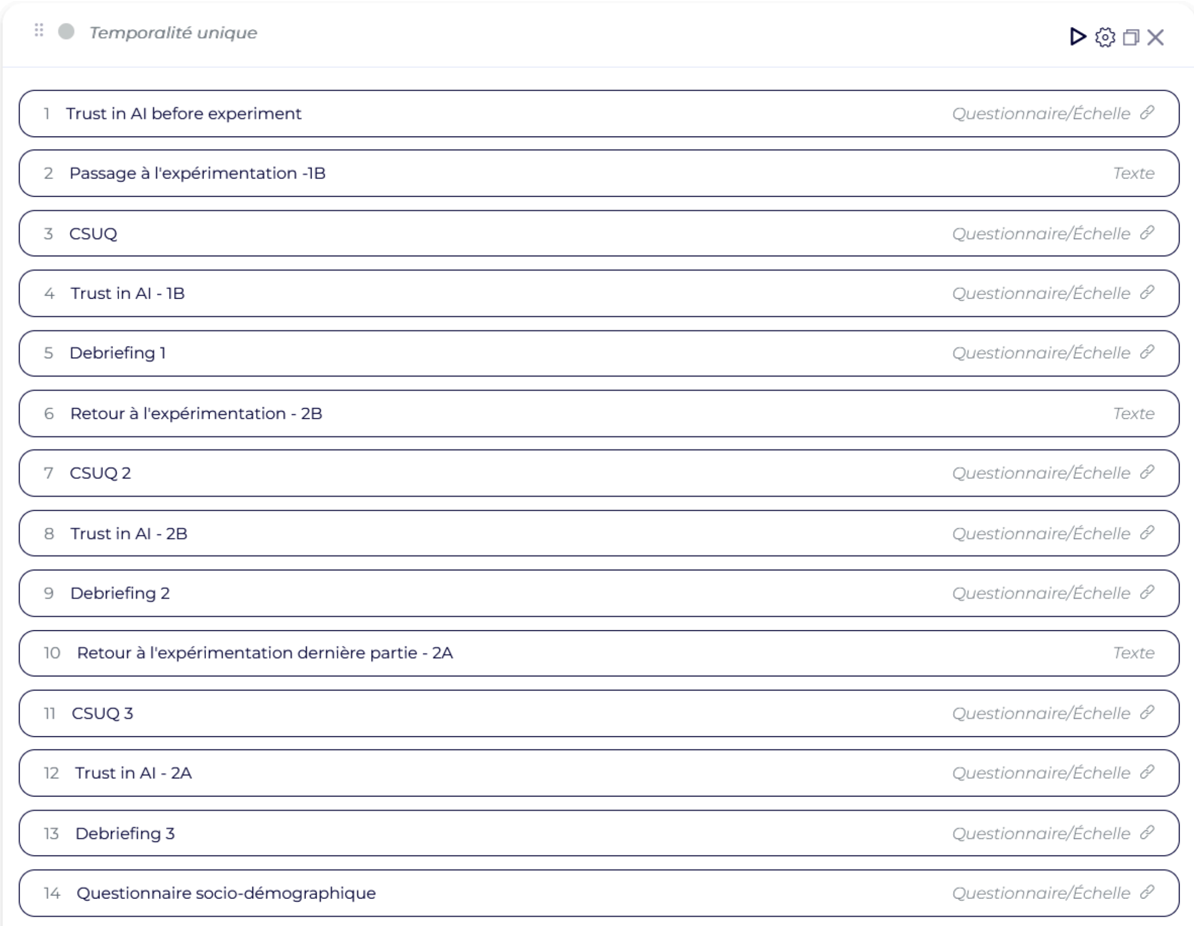
Questionnaires used

The two questionnaires Trust and Computer System Usability Questionnaire (CSUQ) were responded to on the Peac²h platform developed by CATIE. Peac²h is a turnkey web-based solution offering simple, automated online support to help design the evaluation of tools and services. Peac²h includes the evaluation protocols (see Figure 12), the questionnaires and supports in analysing and interpreting the data collected during the experiment, thanks to ready-to-use toolboxes. The two questionnaires can be found in UC#2 Appendix.

After answering the first questionnaire and after each end of a flight scenario the next one was presented to the pilot like described below.

The flight scenario

To give the pilots a concrete example of how to use the assistant, we've created a story with several points where they had to evaluate the options presented to them and to make a choice. Because we're still in the early design phase, and the system isn't yet operational, in this mock-up version the choice of pilots has no bearing on the rest of the story.



There are 3 Scenarios, always from Marseille to Munich, only weather conditions changed (autumn, winter, and summer). For each Scenario, the story is supported by a different type of assistance as mentioned before (1B, 2A, 2B). For each scenario, 2 events took place one after the other:

1. Rerouting: Due to bad weather on the original flight path, a rerouting is necessary (keep the same airport, changing path and the STAR, approach and RWY)
2. Diversion: Around 20nm from Munich airport, ATC inform the pilot, that the airport is closed due to weather conditions (original destination not available anymore, choose an alternate airport)

The original flight plan presented to the pilots is here below (Figure 13) with the associated intentions figures (blue bars in the Figure 14):

MARSEILLE/PROVENCE-MUNICH

LFML/MRS
TakeOff Scheduled 0640
TakeOff 0630

EDDM/MUC
Landing scheduled 0757
ETA 0747

ALTN EDDF 1379 0044

Destination:
EDDM/MUC MUNICH
TAF EDDM 080600Z 0806/0812 11002KT 6000 SCT042
TEMPO 0806/0812 3000 +FZRA BR

DISP RMKS ENROUTE STRONG WIND EXPECTED AND LIGHT TURBULENCES

TAF EDDF 080600Z 0806/0812 20015KT 9999 FEW009 BKN011 OVC014 TEMPO 0806/0812 3000-RA BR

Additional info
1st flight of the day

- PAX max flying back home after holidays 90% of the passengers have checked baggage.
- Fuel in Marseille is cheap.
- Few passengers are in transit in EDDM

Next flight destination: Milan (LIN) LIML

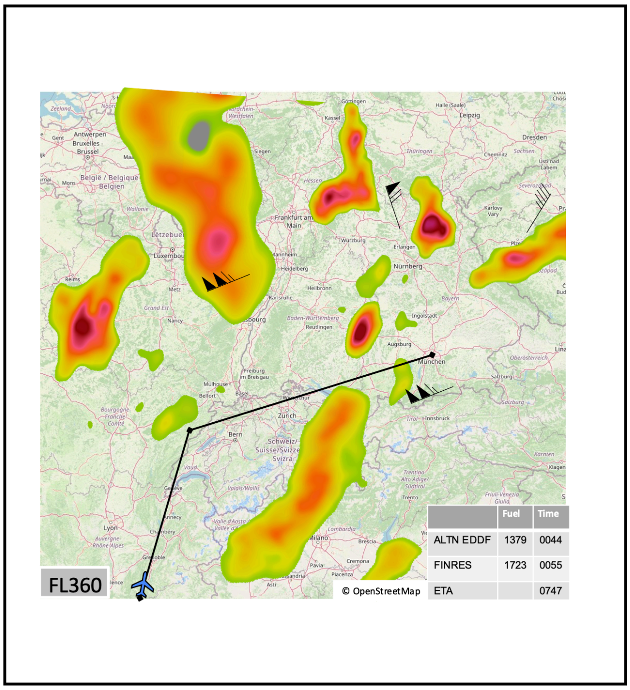
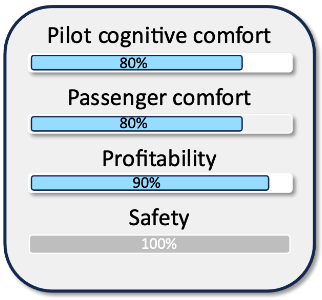


Figure 13. 1st slide to present the flight plan from Marseille to Munich, conditions and additional information.



Pilot cog comfort: On time, known airport destination, bad weather forecast at destination

Passenger Comfort : Turbulence forecasted on route and bad weather forecast at destination

Profitability: On Time, Flight full, with 90% reserved bagages, wide margin FINRES

START

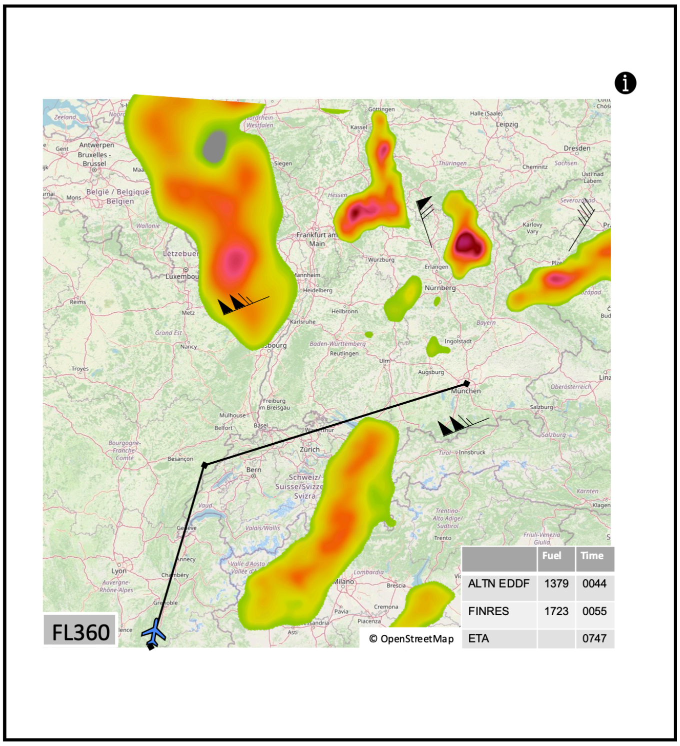


Figure 14. Slide presenting the figures and their explanation for the original flight plan to Munich.

Interfaces

After taking note of the flight plan and conditions, pilots had to select an intention in the interface (cf. Figure 15).

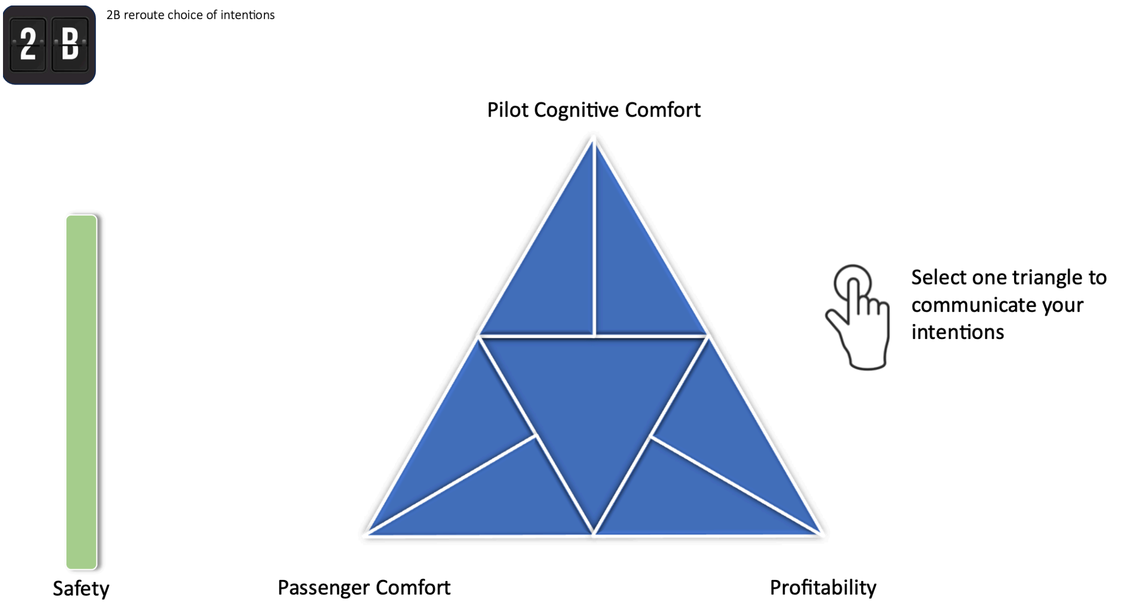


Figure 15. Present the interface of COMBI with all the triangles to select the intentions.

Once the intention is selected, according to the level of the assistant one of the following interfaces was presented to the pilots 1B (Figure 16), 2A (Figure 21) and 2B (Figure 23). In the case of this mock-up, with the limitations of not having the AI to generate the routes and associated scores, only three of the six triangles were clickable. But each time the pilot said aloud the intention, which was noted, and clicked on the triangle next to it.

1B assistant interfaces

In 1B, after pushing the “inflight reroute” button three different maps were presented in the interface (see Figure 17 and Figure 18) according to the intention selected. The first one highlighted in green is the best solution according to the assistant.

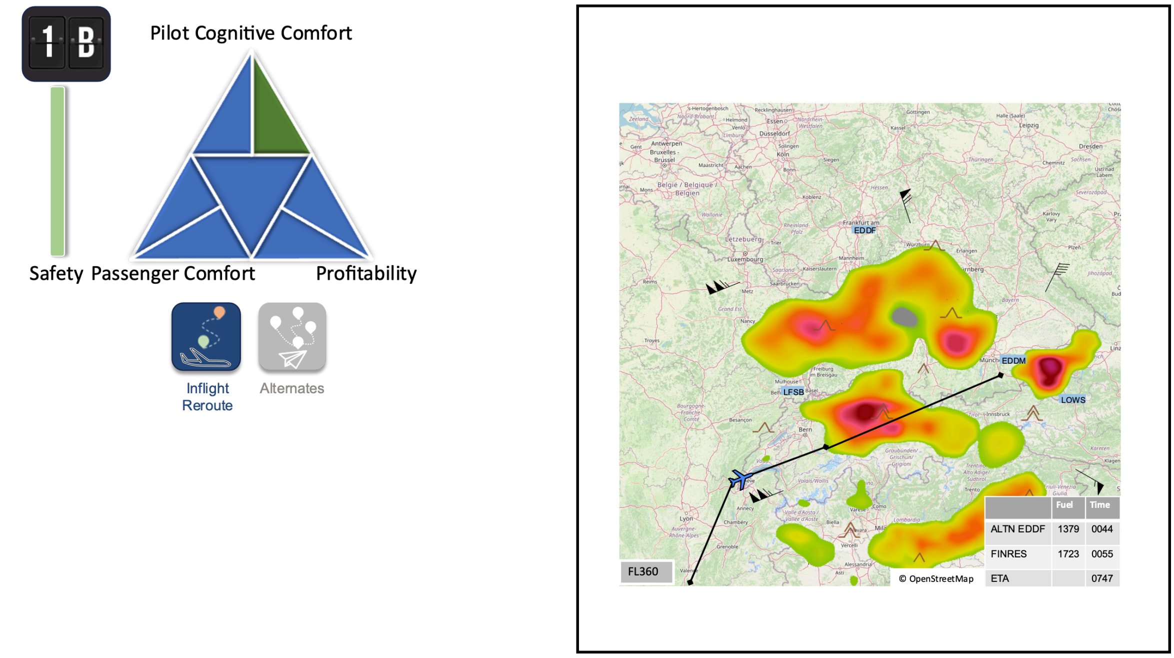


Figure 16. Interface of assistance level 1B on request, showing the new weather condition and proposing the reroute option button. The green triangle represents the selected intention by the pilot.

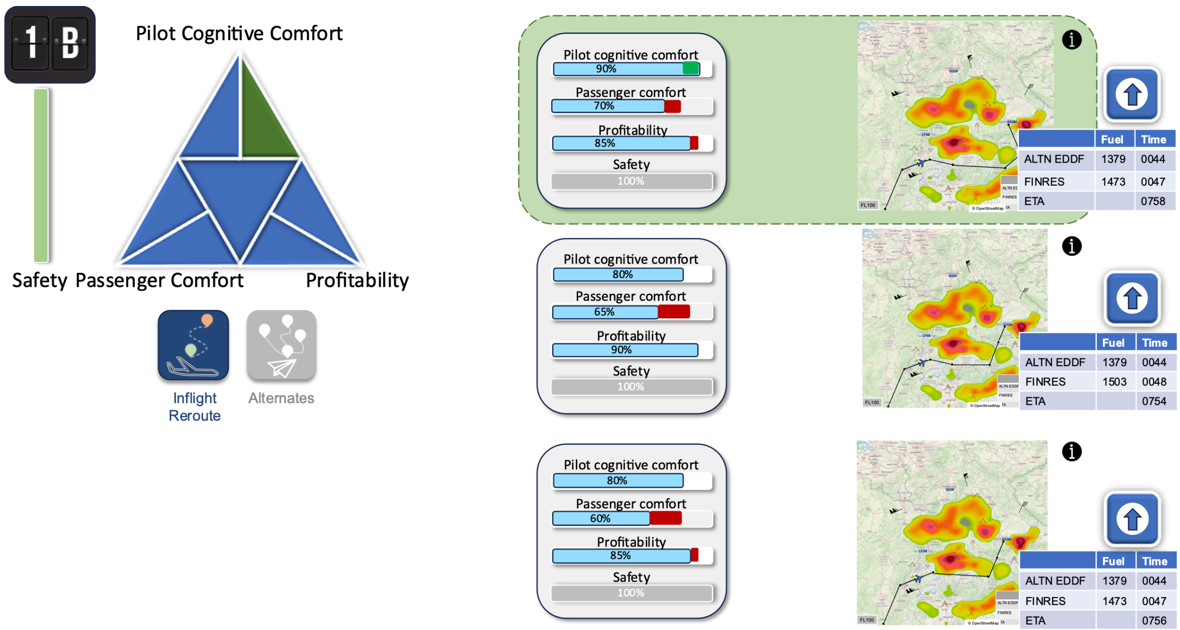


Figure 17. Interface of level 1B after requesting reroute, showing 3 available flights reroute with their respective figures. The blue buttons next to the charts allow uploading the respective flight plan into the FMS directly.

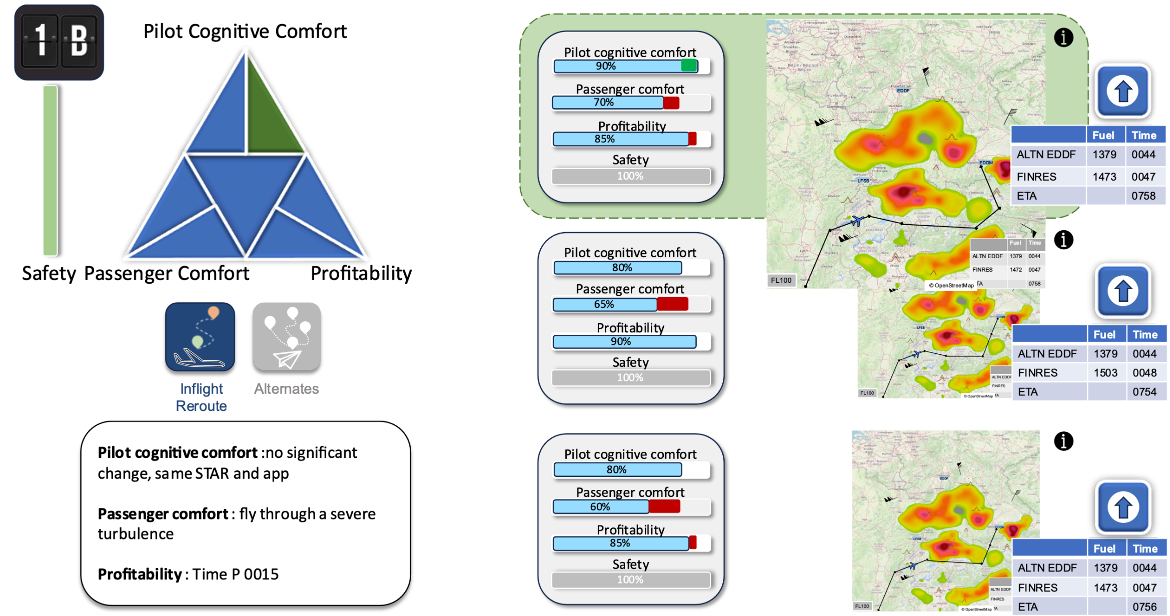


Figure 18. Interface of level 1B showing 3 available flights reroute and the explanation for the first solution, the associated chart is enlarged, after clicking on the "i" button.

The 1B assistant works the same way for the choice of an alternate, after clicking on the Alternates button (Figure 19), three alternate airports were proposed according to the selected intention (Figure 20).

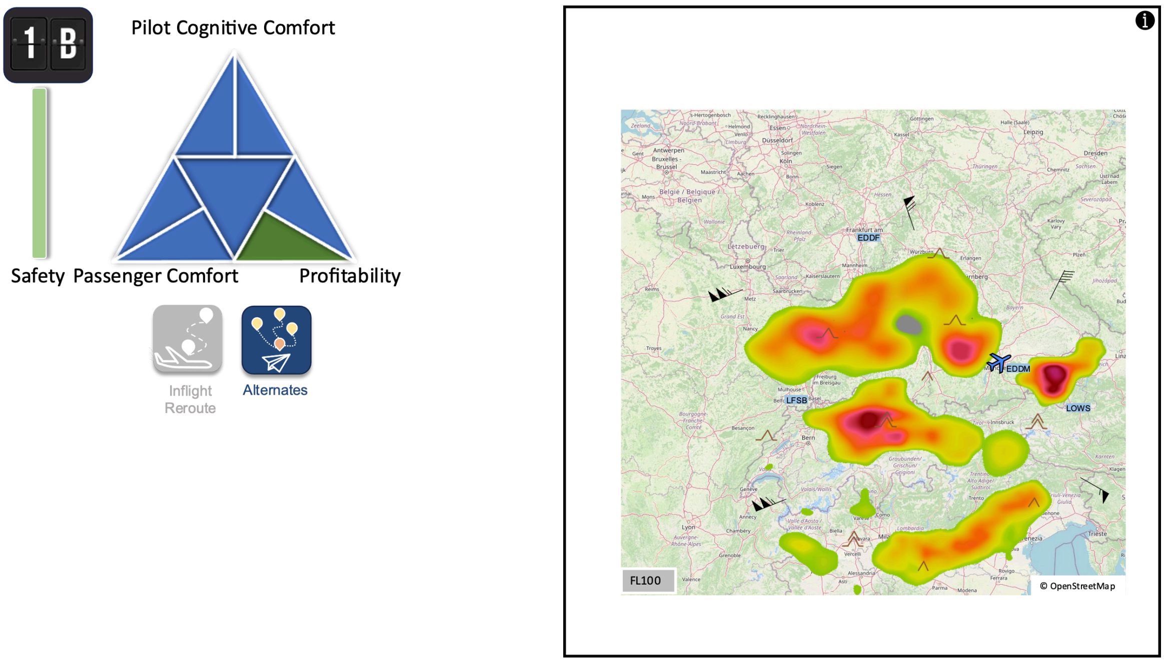


Figure 19. Interface of assistant level 1B on request, proposing the "Alternates" button and the green triangle showing the selected intention.

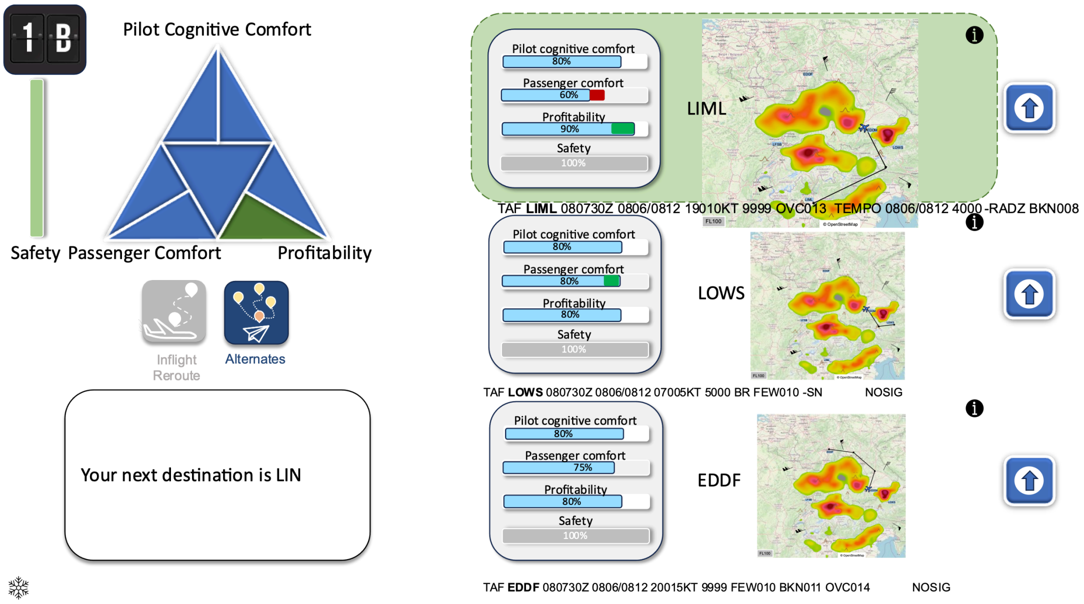


Figure 20. Interface of level 1B showing 3 available flights alternates and the explanation for the first solution, the associated chart is enlarged after clicking on the "i" button.

2A assistant interfaces

In the 2A assistant, once the weather threat is detected by the assistant, a new route is automatically proposed according to the selected intention (Figure 21).

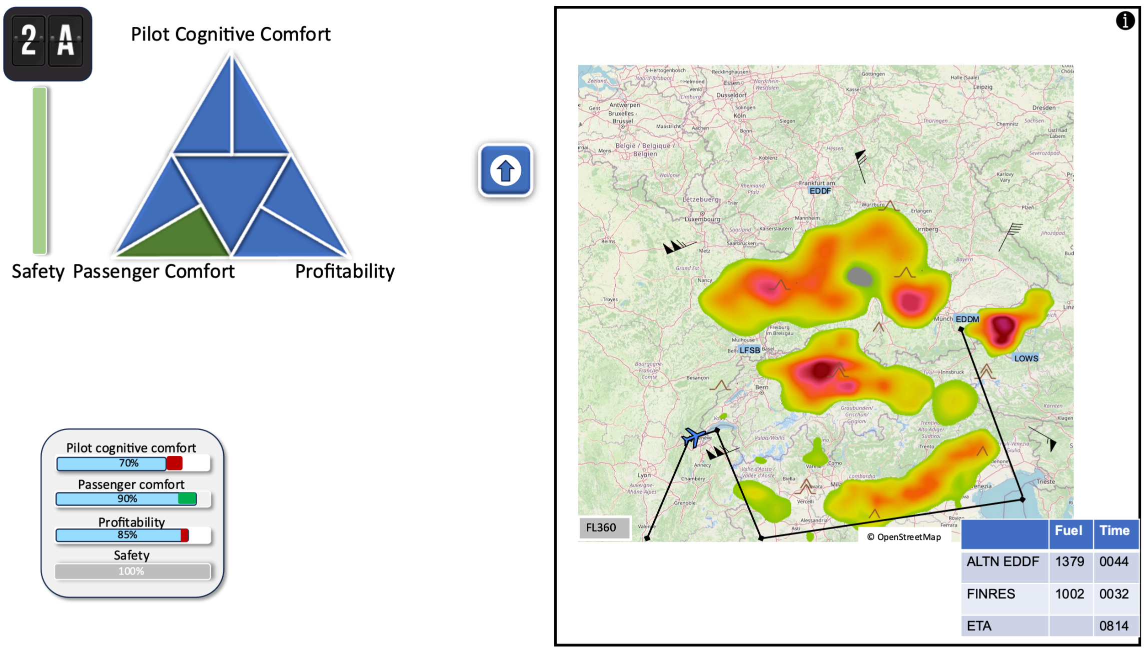


Figure 21. Interface of Level 2A assistance, showing automatically a reroute plan, the figures, the explanation after clicking on the "i" button, and the green triangle for the selected intention.

As for the reroute, once alerted that the original destination airport is closed to traffic, the assistant proposes automatically an alternate airport to the pilot (Figure 22).

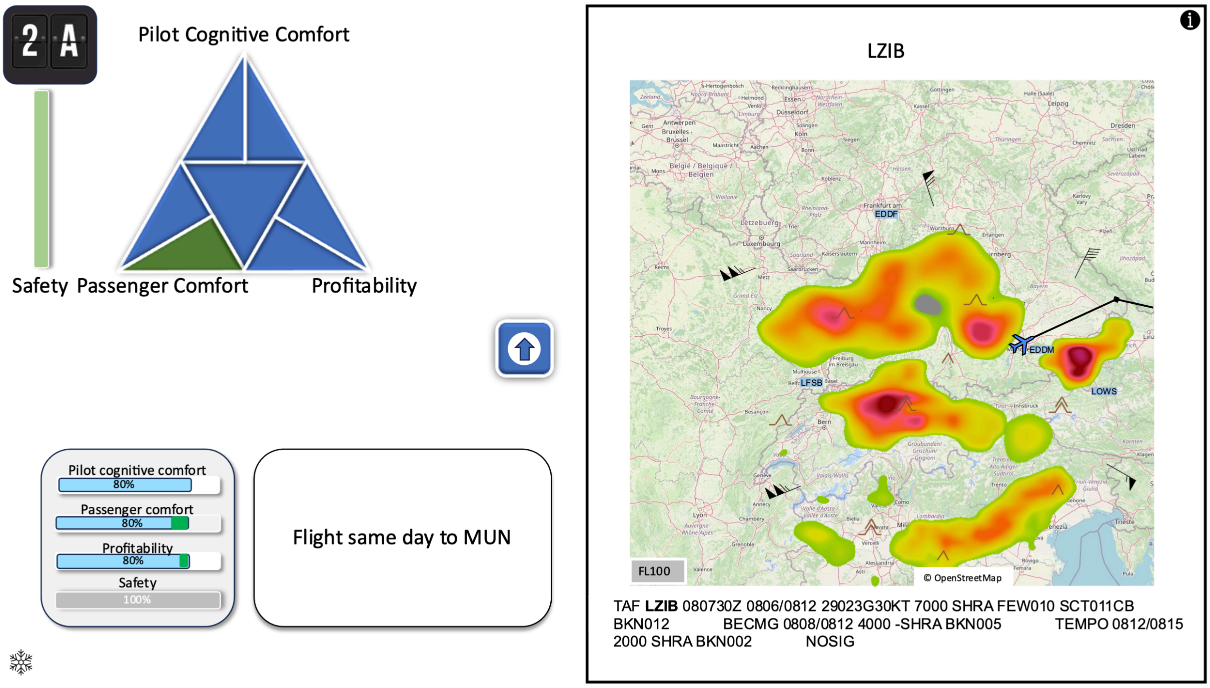


Figure 22. Interface of Level 2A assistance, showing automatically an alternate airport, the figures, the explanation after clicking on the "i" button, and the green triangle for the selected intention.

2B assistant interfaces

In level 2B, the assistant automatically proposes a new route after detecting the weather threat and by default the "least negative impact" option (cf. Figure 23). The pilot can choose to change for "best compromise" button option, proposing another route or another alternate in the case of the alternate step (cf. Figure 25 here for the alternate choice with "Best compromise" button selected).

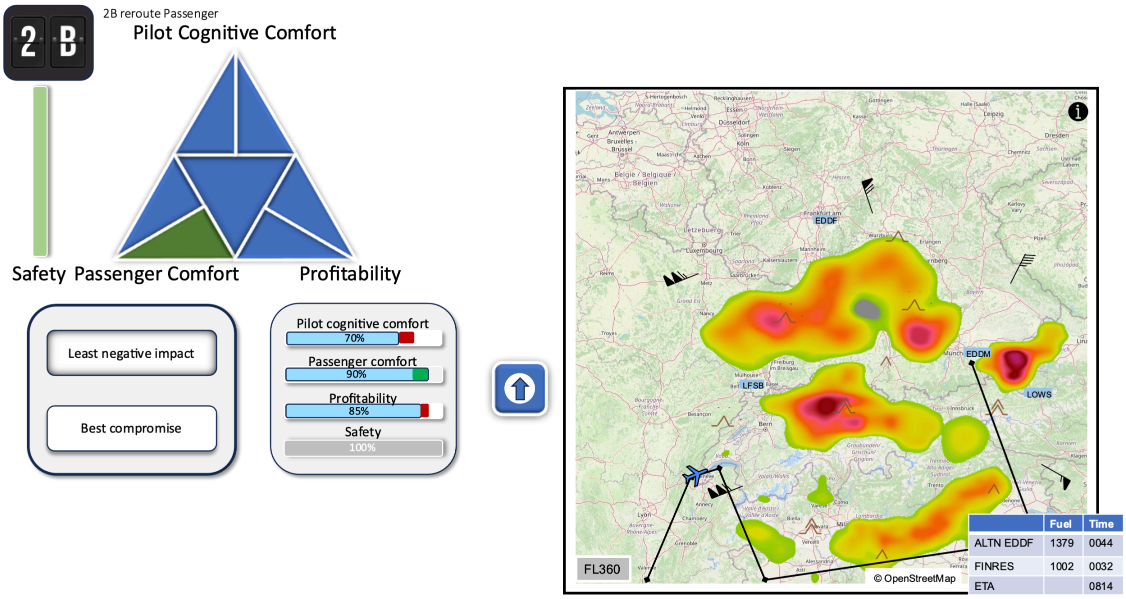


Figure 23. Interface of Level 2B assistance, showing automatically a reroute plan, with the least negative impact solution selected, the figures, and the green triangle for the selected intention. A button "Best compromise" can be selected.

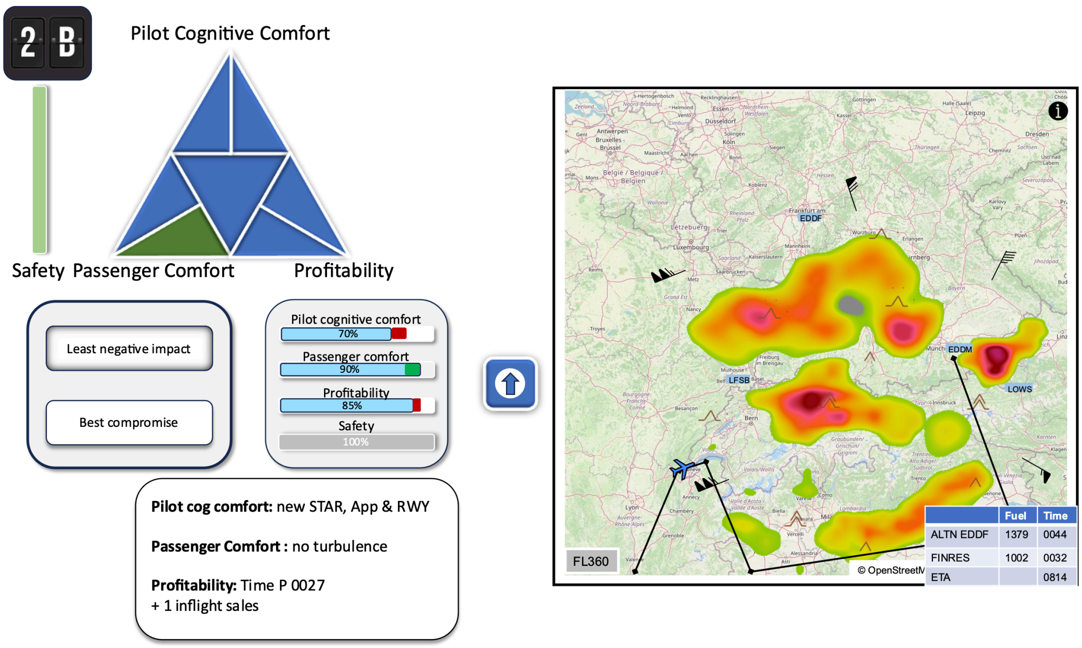


Figure 24. Interface of Level 2B assistance, with the explanation after clicking on the "i" button.

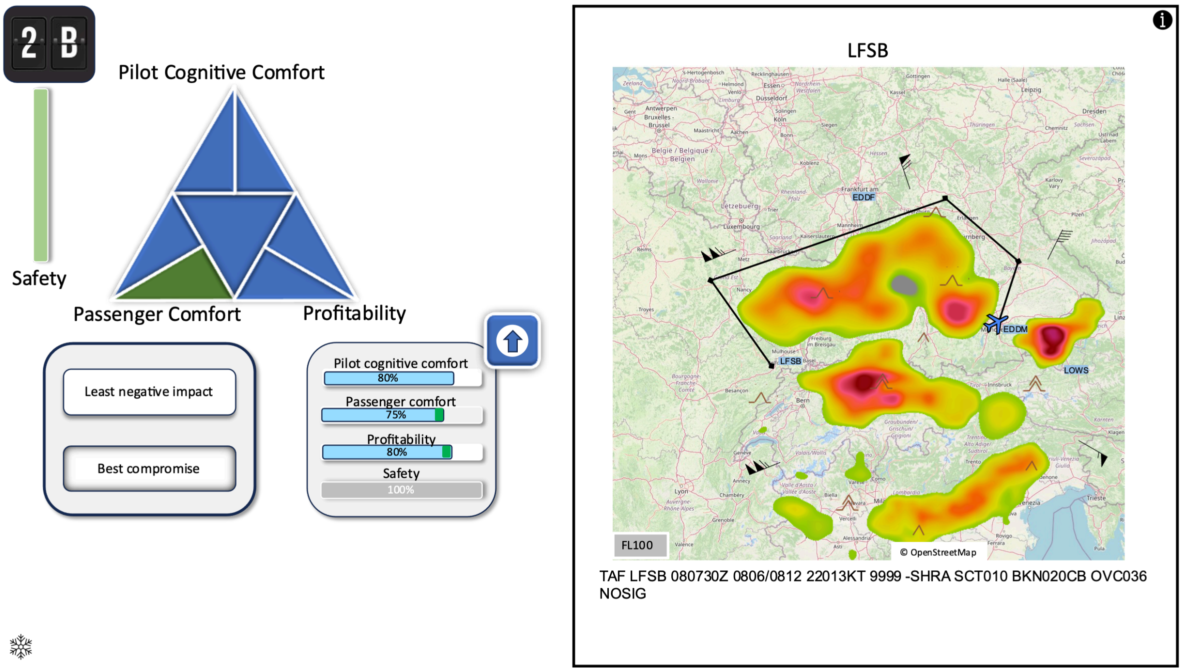


Figure 25: Interface of Level 2B assistance, with the "Best compromise" selected, the figures, and the green triangle for the selected intention.

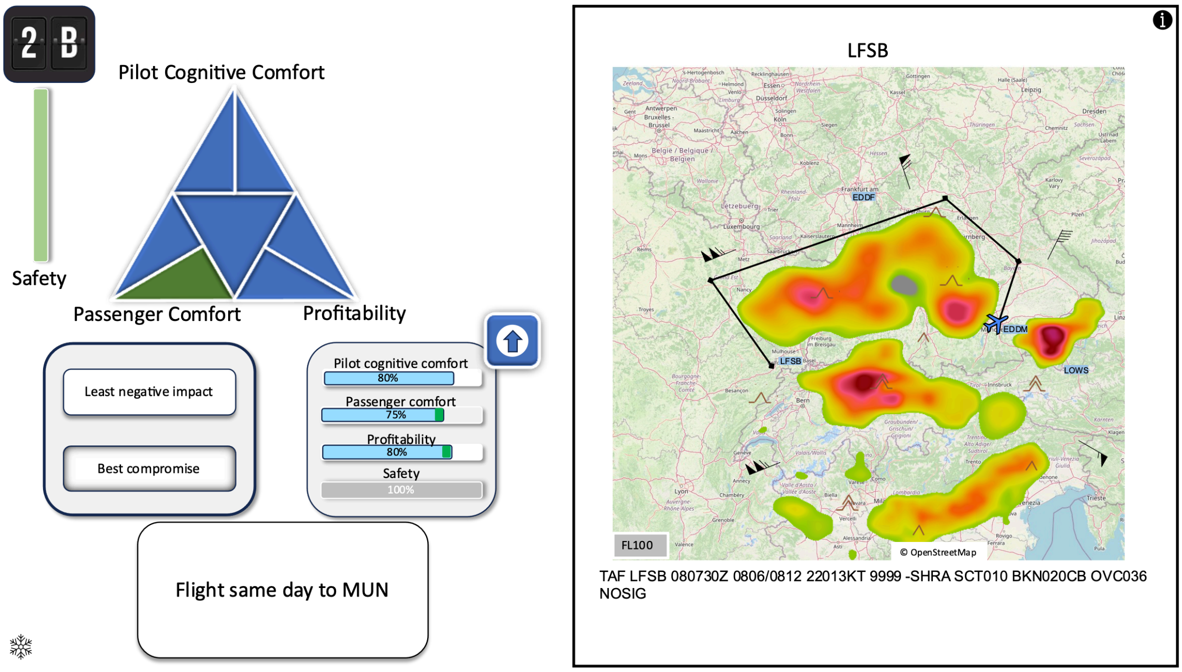


Figure 26: Interface of Level 2B assistance, with the "Best compromise" selected, the figures, the explanation after clicking on the "i" button, and the green triangle for the selected intention.

Debriefing

The pilots were asked to think out aloud, to allow an understanding of their thinking process, decision-making and the information they relied on. In order to obtain more information on the level of the tested assistant, a brief debriefing took place each time to allow the participant to give his/her opinion on the assistant and the interface. In addition, after each end of flight scenario for an assistant level, two questions were asked to the pilots:

1. On a scale of 1 to 10, how useful is this assistant against being alone in the cockpit?
2. What aspects of the assistant do you think would contribute the most to its adoption?

At the conclusion of the three levels, participants engaged in a debriefing session to share their thoughts on the assistance provided across these levels. This provided an opportunity for us to inquire about the prioritisation of information, distinguishing between essential content to be directly displayed and supplementary details available upon request. Furthermore, we gathered feedback on the additional information needed for informed decision-making, as well as insights on interface design and user interactions.

Results method overview

To validate the three objectives of UC2, an experimental protocol was developed. Participants were asked to evaluate three interface modalities representing different types of assistance: decision support, cooperation, and collaboration. Before the tests commenced, each participant completed a questionnaire assessing their trust in AI. After each evaluation session, participants again completed the AI trust questionnaire, as well as an assessment of the usability (CSUQ) of the tested interface.

The results of the experiment were analysed to address the specific objectives of UC2. Objective UC2-OBJ-02 aimed to identify the key features of each type of assistance (decision support, cooperation, collaboration) that promote teamwork requirements assurance and effectiveness. To meet this objective, we rely on the analysis of verbatims. Objective UC2-OBJ-04 and UC2-OBJ-05 examined whether cooperation and collaboration improved the decision-making process compared to decision support alone, and whether collaboration outperformed cooperation in in-flight rerouting situations.

The experiment results provided insights into how each assistance modality affects pilots' trust in AI, as well as their perception of interface usability. Additionally, the data collected allowed for the evaluation of the performance of different assistance modalities in terms of decision-making process efficiency and teamwork requirements assurance.

4.3 UC#2 TRL

System	Current TRL	Evidence	Target TRL	Evidence
COMBI	3	Former trials with the French Air Force for defence context (TRL 4). Degradation of TRL due to the need to adapt for the commercial flight domain (Hourlier, 2022).	5	Evaluated in real time flight simulator in an operational representative flight scenario for diversion (commercial flight domain). Data sources synthetic.

The UC2 system is based on a single module that employs artificial intelligence (AI) technology.

The COMBI prototype has already been tested in a military context. Indeed, the same design methodology is being employed in the development of the civilian version of COMBI. It is possible that specific components may be reused, such as particular interface elements or the foundation of specific trajectory calculation algorithms. However, these components will also require adaptation to align with civil domain standards. Moreover, additional components will be required to address the operational challenges inherent to the civil sector. It is necessary to recreate the top-down (human to machine) and bottom-up (machine to human) translation bricks in their entirety, based on the existing methodology. For this reason, the current TRL level is 3, which represents a reduction compared with the TRL 5 military prototype.. This reduction is due to the change in the field of application.

In VAL1, the interface, communication and recommendation principles employed in the military version are utilised, albeit with adaptations to align with the new context. The system version in VAL1 is entirely fictitious, and the outcomes are used as a foundation for developing the interface and algorithms.

With regard to the communication module, the new intentions defined during interviews with the pilots for WP3 are being taken as a point of reference. Should the experiments yield conclusive results, these intentions will be incorporated into VAL2.

4.4 UC#2 Results

Preliminary questions

Questions descriptions

First and foremost, we formulated two questions to answer objectives UC2-OBJ-04, UC2-OBJ-05 and UC2-OBJ-02. Interpretation of the responses gives us the participants' views on the usefulness of the assistant compared with using the cockpit alone and the aspects contributing to its adoption.

For the question "On a scale of 1 to 10, how useful is this assistant against being alone in the cockpit?" (UC2-OBJ-04 & UC2-OBJ-05 of validation plan), participants' responses provide valuable feedback on the perceived utility of the assistant in real-world aviation scenarios.

Higher ratings indicate that the assistant is seen as more beneficial compared to operating alone in the cockpit, suggesting that the assistance features are effective in enhancing decision-making and overall flight operations. Conversely, lower ratings may indicate areas where the assistant falls short in meeting pilots' expectations and requirements for in-flight assistance.

Regarding the question "What aspects of the assistant do you think would contribute the most to its adoption?" (UC2-OBJ-02 preliminary assessment), participants' responses shed light on the key features and functionalities of the assistant that are most valued by pilots. Common themes and patterns in participant responses can highlight aspects such as user-friendliness, accuracy of information provided, integration with existing cockpit systems, and adaptability to different flight scenarios. Understanding which aspects contribute most to the adoption of the assistant informs future development and refinement efforts to enhance its effectiveness and acceptance among pilots.

In summary, interpreting the results of these questions involves synthesising participant feedback to identify strengths and weaknesses of the assistant, as well as key factors driving its perceived usefulness and adoption. This analysis can guide iterative improvements and optimizations to ensure that the assistant meets the evolving needs and expectations of pilots in the aviation environment.

The next section describes the results of these two questions.

Usefulness of the 3 assistants

Interpretation of the results for question 1, distinguishing between the 5 situations and the 3 types of assistants (1B - decision support, 2A - Cooperation, 2B - collaboration):

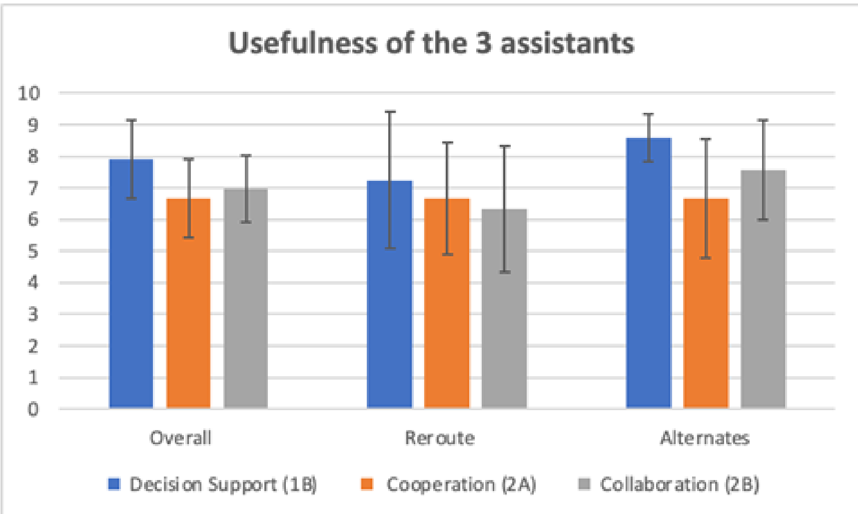


Figure 27: Overall usefulness of the 3 assistants, only for reroute, only for alternates (diversion)

Situation	Decision Support (1B)	Cooperation (2A)	Collaboration (2B)
Overall	7,92	6,67	6,96
Reroute	7,25	6,67	6,33
Alternates	8,58	6,67	7,58

Table 9. Mean score of each assistant, overall and in specific situations (Reroute, Alternates).

The non-parametric statistical tests (Kruskal-Wallis test) carried out on these results do not allow us to conclude that there is any significant difference between the assistance for all the conditions.

These results suggest that overall, participants tend to prefer the 1B type assistant (decision support), followed by the 2B type assistant (collaboration), while the 2A type assistant (cooperation) is less favourably rated. However, it is important to note that preferences may vary depending on specific operational situations.
Nevertheless, we can conclude that any of these assistants would be useful (score above 5), for any situation (Overall, Reroute, Alternates).

Trust Questionnaire

This questionnaire assesses individuals' tendencies to trust technology in various contexts. It can be used to understand participants' attitudes and perceptions regarding the reliability, usability, and effectiveness of technology in supporting their tasks and decision-making processes.

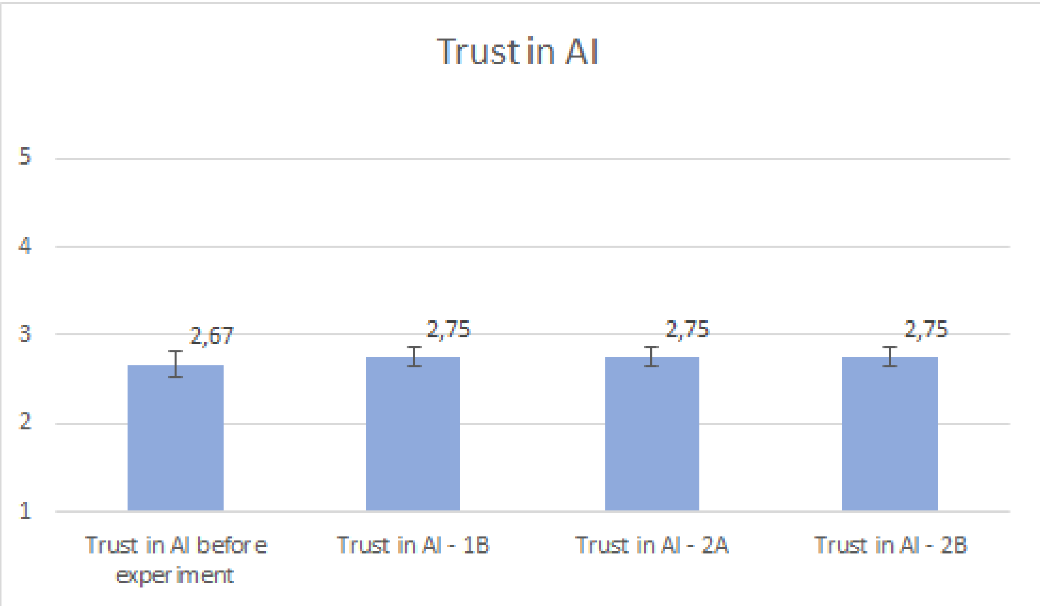


Figure 28: Trust in AI before experiment, in 1B, 2A and 2B

Based on the Kruskal-Wallis test results, it appears that the experience did not lead to significant changes in the subjects' overall trust in AI across the different conditions (1B, 2A, and 2B). The p-value obtained from the test (0.8) shows that there is not enough evidence to conclude that the trust scores significantly varied among the conditions after the experiment.

CSUQ (Computer System Usability Questionnaire)

The Computer System Usability Questionnaire (CSUQ) was created by Lewis and Sauro in the 1990s. It was developed to assess the usability of computer systems in a wide range of contexts, including software development, user interface design, and user experience research. Lewis and Sauro were working at the IBM corporation at that time, drawing on their expertise in human-computer interaction and user experience research gained from their work. Based on principles of psychometrics and ergonomics research, the CSUQ comprises multiple subscales evaluating various aspects of usability. Respondents provide ratings on a Likert scale (1 to 6), enabling researchers and designers to understand user perceptions and identify areas for improvement to optimise the user experience.

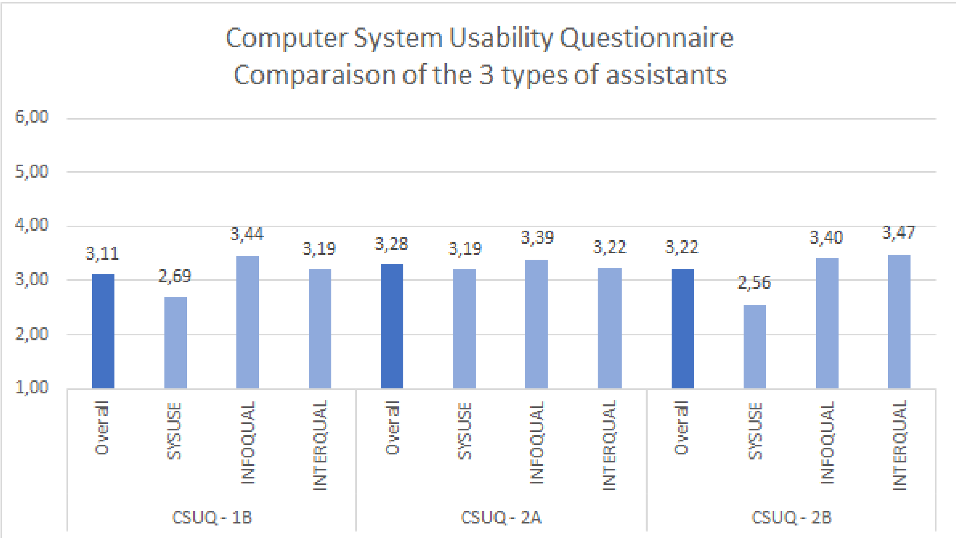


Figure 29: Results of the CSUQ questionnaire for 1B, 2A and 2B intelligent assistants, and by dimensions, SYSUSE (System Usability), INFOQUAL (Information Quality), INTERQUAL (Interaction Quality).

In summary, cooperation (2A) seems to offer an overall improvement in the user experience compared to decision support (1B) and collaboration (2B). The system usability seems to be better in 2A. The interaction quality seems to be better in 2B.

Key feature features elicitation and evaluation (OBJ-2)

UC2-OBJ-02: What are the key features for each type of assistance (decision support, cooperative, collaborative) that enable teamwork requirements assurance and effectiveness?

The assistants 1B, 2A and 2B were initially envisaged as variants to provide, respectively, decision support, cooperative Human-AI teaming and collaborative Human-AI teaming, according to the definitions in the EASA Concept Paper. Nevertheless, it was not evident to the pilots that the different implementations implied different roles of the assistant (and consequently of the pilot). The captured perception is that all variants were considered as decision-support tools, with no delegated authority to perform any tasks automatically, no capability to discuss with the pilot the best course of action to apply, and no capabilities of dynamic task sharing. This is due to the choice made to communicate by high level intentions in all three levels, and to the fact that the VAL1 is not done on a simulator and without the AI algorithms, but only as a mock-up, with the technical and immersion limits that this represents.

The key features of interest captured from the pilots for the role of decision support, were:

- **The need of an IA is justified by the complexity of the situation.** The less complex scenario was considered easy to manage by the pilots. Finding a new airport and a new flight plan is a complex situation where the IA is very useful.
- **Intentions assessment must be substantiated with complementary information** (e.g., fuel margins, ETA, performance calculations, weather/traffic forecast) to align understanding and support decision-making;
- **Value is given to the simultaneous presentation of (key aspects of) several solutions**, enabling a faster understanding of the trade-offs; This consideration should be applied to any role of the assistant. 3 solutions look like a good quantity to compare and make a quick decision.
- **Value is given to the possibility of exploring a broad solution space**, mostly considering the diversion scenario (by having more presented options, being able to select other airports…);
- **Value is given to information quality**: recency, expected updating frequency and confidence levels;
- **Value is given to the information panel (explainability of the solution)** to understand the proposal of the system and align the mental model about intentions.
- **Value was given to further abstracting intentions trade-offs.** The solution labels presented in assistant 2B were considered useful by some of the pilots - they could represent an assessment in a higher level of abstraction, if good cognitive alignment about their meaning is reached.

Some glimpses were captured indicating that pilots would be uncomfortable with the system directing them towards one single solution (which would be the task envisaged for assistant 2A), instead of asking them to choose among the possibilities. Scepticism was also shown about the candidate capabilities of a collaborative assistant (2B), such as:

- performing context-relevant cross-checking of the pilots' considerations in the decision-making process;
- providing relevant suggestions on intentions redefinition.

The possibility to query the system for information and assessment considerations in a conversational fashion was pointed out as valuable by one of the pilots.

Also, some pilots said that automatic detection of deviations from initial flight plan assumptions (e.g., change of winds, weather forecast, etc), with subsequent presentation of strategies and outcomes would be of value.

The following sub-sections present insights on these aspects:

- The three intentions the associated requirements
- Information needed to support decision-making for reroute and diversion
- HAT functions proposed in VAL1 and new ones based on pilots' feedbacks

- Operational explanation (OpXAI)
- HAT insights/requirements
- HMI

Intentions

Intentions act as a means of communication and context alignment in high abstraction levels. To accelerate such an alignment and thus contribute to the decision timeliness and quality, these intentions must correctly capture the meaningful aspects that may drive the pilots' decisions among a set of feasible solutions.

Overall appreciation:

- The set of intentions presented was considered representative of the most relevant aspects taken into account in pilots' decision-making processes. On the other hand, this could be enriched by a specific sub-set of intentions on "pilot cognitive comfort";
- Pilots recognized the value of intentions to accelerate communication.
- Operating the decision-making ONLY on the intention trade-off was not possible. Pilots needed extra information for decision-making.
- Intention cognitive alignment was shown to be critical for pilot trust. Misalignment in expected intent evaluation affected fluidity of the decision process. Intention cognitive alignment can be defined here as the consistency between the mental model of the end user and the IA model when assessing how the operational intentions relate to the set of technical parameters associated with a proposed solution.
- After having defined the intentions, pilots start the evaluation of the solutions by a safety analysis. It is mainly associated with the very complex meteorological situation.
- Some pilots tried to reverse-engineer the intentions assessment to infer the assistant's considerations and calculations, in order to cross-check them. The lack of clarity on which information was taken into account and how the intention assessment was performed seemed to be key for this behaviour;
- The pilots did not have a clear picture of what would be the best means to present the plethora of information related to each intention:
    - One pilot showed concerns on the system trying to guess what information was needed and not being able to do so, given the highly contextual nature of the task.
    - One pilot was concerned about the workload associated with an active search for information/explanations.

PILOT COGNITIVE COMFORT was the most salient intention in the decision process by far, due to its direct impact in assuring flight safety. It is clear that this intention integrates information relative to multiple threats, providing a notion of the level of challenge associated with the mission accomplishment for a selected option. But assumptions regarding the underlying model to evaluate pilot cognitive comfort varied among pilots.

The possibility of integrating information about pilot states (e.g., fatigue, attention levels) and previous experience (e.g., familiarity with region and specific airport) to this assessment was investigated during the interviews. Pilots' assessments were not uniform considering whether such individual tailoring would be beneficial or not to the cognitive alignment.

REQUIREMENT: knowledge of the "pilot cognitive comfort" assessment model must be clear to the pilot.

PROFITABILITY assessment must integrate lots of information that may be considered secondary to the pilot, as they do not relate directly to safety. This includes the impact of missing connections and crew duty periods extensions caused by a delay, the availability and cost of services in an alternative destination, transport and accommodation options available for passengers and crew etc. The overall perception is that the direct assessment of this intention is quite valuable as it reduces the need of alignment with airport operation control centres or consultations to guidelines provided by the airline.

Some pilots expressed concerns on whether all this information was captured and if the assessments were coherent. Nevertheless, those concerns were shown less often than those related to the Pilot Cognitive Comfort assessment.

One pilot expressed that it would be valuable to rank the options by profitability, having a clear view of the airline preferences, and then select the first option that would be considered adequate in terms of the other intentions in the given context.

REQUIREMENT: the profitability assessment should represent a clear and updated view of the airline preferences regarding its commercial results.

PASSENGER COMFORT is considered a relevant intention, but in general does not take many factors into account. The pilots' seemed less concerned in having a clear picture on how this intention was assessed and didn't point out any inconsistencies about that in the experiments. This intention is less important in very complex situations as presented in the test. However, pilots consider that it is important in less complex situations.

Information needed to support decision making:

From the debriefing sessions a set of information types needed to support the decision-making process in route replanning was collected.

Some of them are expected to be presented immediately, with the solution proposals, in order to characterise them, improve cognitive alignment and accelerate the trade-off:

- Route characterization on map, showing at least the 2D path to the destination;
- ETA and/or remaining flight time;
- Fuel margins (to destination and to alternate, after reaching destination);
- Forecasted weather conditions, mainly in the destination, at the considered time of arrival.

Other types of information are expected to be available for consultation, allowing the pilot to verify the cognitive alignment or dig into potentially relevant details:

- Performance parameters and calculations;
- Approach type and terrain details (e.g. RWYs heading and length);
- Forecasted traffic congestion at arrival;
- NOTAMs, ATIS, METAR, TAF;
- Destination services/infrastructure and associated costs/airline contracts;

- Transfer means for passengers and crew from alternate airport;
- Data accuracy and recency;
- More detailed considerations on weather evolution over time including clouds height, possibly historical patterns;
- Justification on intentions assessment in terms of technical parameters.

RECOMMENDATION: integrate, as possible, the information mentioned, namely that mentioned to be presented immediately with the solution proposal.

Functions

Some functions/capabilities for the assistant mentioned by the pilots as interesting are:

- Ability to automatically detect relevant changes affecting route planning and alert the pilot;
- Ability to translate the intentions set by the pilot into technical parameters, to improve prioritisation of the possible route re-planning solutions.
- Ability to account for the weather dynamics (forecast) into the solution proposal;
- Ability to integrate information from different sources, including outside world (e.g., internet weather forecast);
- Ability to provide historical trends (e.g., weather and traffic dynamics, accounting for seasonal or daily patterns, etc);
- Ability to integrate information on the pilots' experience, cognitive state or preferences into the assessment of pilot cognitive comfort;
- Ability to account for the information mentioned as relevant for decision-making, as much as possible, to perform the assessments of intentions met by each solution;
- Ability to timely justify the assessment of intentions and safety/feasibility of multiple proposed solutions, at an adequate level of abstraction;
- Ability to provide further information supporting the assessment of intentions and safety/feasibility;
- Ability to provide awareness of potential degradation of proposals validity, due to data quality aspects;
- Ability to inform the AOCC of decisions;
- Ability to load final decisions to FMS.

OpXAI

The operational explanation was provided in three layers:

- An overview of the assistant concept, capabilities and interfaces.
- Immediate information during the operational exercise, provided mostly in a graphical format, enabling the pilot to interpret most of key elements of the situation and the proposed solution: a 2D representation of the route path, waypoints, airport, terrain information, and meteorological (static) information. Other key information that characterised the proposed solution was presented at the same time: time of arrival, fuel margins, TAF and the high-level assessment of the outcomes in terms of intentions, which are compared to a reference point.
- On demand, the pilots could reach a limited amount of complementary information that was related to the intentions assessment.

As the pilots were exposed to a very limited number of scenarios, after a short familiarisation with the assistant concept and capabilities, it is difficult to conclude if the initial set of information would fill the pilots' explanation needs in a typical re-route scenario, or if further information would always be required. Nevertheless, the pilots indicate that having a certain amount of information provided at the same time for different solutions, enabling a quick comparison, would be beneficial. This was pointed out as the main reason for preferring the 1B variant, as it showed more options at once.

The complementary information was always consulted, and most of the pilots would request even more details, if possible. Although pilots found it positive to be able to access all the information that could contribute to the decision integrated and organised in the assistant, the best way to organise the information was not evident.

These three layers of operational explainability – latent (in a training material), immediate (quickly characterising the proposed solution with respect to the situation), and complementary (which could be reached with different interaction approaches) – are paramount to assure effectiveness of the assistant concept. Those should be carefully developed to improve the representativity of experiments in VAL2.

Human AI Teaming insights/requirements

As mentioned, the aspects of Human AI teaming, as expounded in the EASA concept paper, were not directly addressed by the experiments. Nevertheless, observation of user behaviour and elements captured in the debriefing inform teaming related features to be further addressed in follow-up research. Regarding aspects of teaming, some key ideas captured from the experiments and debriefing interviews were:

- Cognitive alignment assurance**.** Intention, as a proposal for accelerating understanding and decision making, is contingent on alignment of intention model between human and machine. Not understanding the underlying model for intention assessment, as well as being presented to evaluations of solutions that were not aligned with own assumptions of such models, was often observed in the experiments, and clearly impaired the decision-making agility.
- **Functional qualities**. some abilities - or qualities - regarding what the assistant can provide establish a foundation for the perception of usefulness and thus, the feasibility of establishing a team. Namely:
    - **Time to provide alternatives**: information and proposals need to be provided in useful time to enable understanding, evaluation, decision and action implementation;
    - **Integration of disparate data sources** and their transformation into information useful for decision
    - **Assessment of confidence levels:** data validity must be evaluated with respect to recency, quality and coverage of the Operational Design Domain (ODD)
- **Boundaries of performance**. Understanding the performance boundaries of the assistance is paramount to build an adequate level of trust in the system This allows the human to delegate responsibilities to the assistant for down-selection of solutions in well-defined, low-impact contexts, or for higher level framing of the solution quality at

more complex situation, while keeping the right level of awareness and readiness to override the system.

HMI

Some further functions/capabilities for the assistant were mentioned of interest by the pilots:

- Ability to interact and manipulate the proposed route. E.g., add geofencing, move trajectory, etc. The assistant would provide evaluation of impact on the change;
- Natural voice interaction;
- Use colour coding to show situation at airport;
- Integrate several information into the solution proposal (see section 0)
- Ability to configure fast access buttons (e.g., quickest route)
- Ensure language codes alignment (e.g., ALTN, FINRES)

4.5 Final discussion and recommendations for VAL2

Regarding the VAL1 objectives, the main insights were:

UC2-OBJ-02: What are the key features for each type of assistance (decision support, cooperative, collaborative) that enable teamwork requirements assurance and effectiveness?

- UC2-OBJ-04 2A variant: HAT cooperative teaming improves decision making process for on air re-route situation vs. decision support assistance
- UC2-OBJ-04 2A variant: 2B variant: HAT collaborative teaming improves decision making process for on air re-route situation vs. HAT cooperative teaming

In our observations, we have noted marked differences in decision-making processes across scenarios 1B, 2A, and 2B. The use of 2A assistant is characterised by rapid decision-making, often due to the absence of alternative options. Conversely, pilots took significantly longer to make decisions with the assistants 2B and 1B.

This disparity in decision-making timelines can be attributed to humans' limited cognitive resources, as highlighted by Rasmussen's SRK (Skill, Rule, Knowledge) model. The more the Human relies on analytical knowledge, the slower and more cognitively demanding the decision-making process becomes. This phenomenon is known as the paradox of choice, where the attempt to avoid missing out on the best option can prolong the decision-making process and create frustration over unchosen options. Furthermore, offering several choices leads to a need for cognitive closure which is the desire of a person to obtain a clear answer in order to avoid uncertainty and regret, resulting in a cognitive cost linked to frustration.

It is also time stress related, in the scenario the pilots had plenty of time to make their decisions. As shown in the figure below, the less time stress there is, the more analytical the decision-making can be, and therefore the longer it takes.

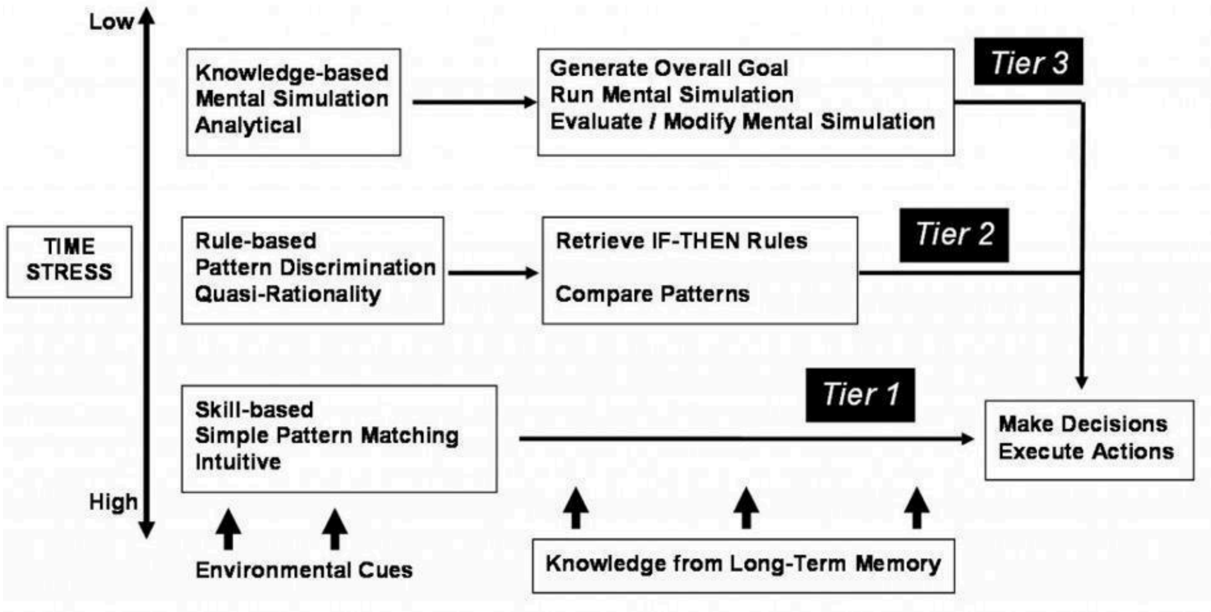


Figure 30: Naturalistic Model of Decision Making (Elgin & Thomas, 2004)

To address this issue, using AI trained algorithms rather than human designed paths in the VAL1 mock-up could help pilots better understand available options and increase their confidence in the VAL2 system, thus reducing the need to offer multiple choices. Additionally, during VAL1 evaluations, pilots requested some additional information necessary to make their decisions, which will be integrated into VAL2 as said in section 4.3.2.

Today's airline pilots are accustomed to reasoning with decision-making methods such as FORDEC, which require thorough evaluation of options before making a decision. Consequently, the presentation of singular solutions by assistants 2A and 2B should align with COMBI's internal FORDEC process. This approach is particularly relevant for alternate choices, which raise heightened safety concerns for which FORDEC was specifically designed. Due to their familiarity with this kind of decision-making process, the adoption of assistants 2A and 2B, as designed for VAL1, may require additional time for acceptance and integration.

Results of VAL1 have provided some recommendations for the final demonstration (Validation 2):

- Each IA concept has positive and negative characteristics. Pilots have identified those for adequate human-IA teaming. For the second validation, it is necessary to implement those requirements to ensure performance and not focus the development with a precise concept of IA as has been proposed by the EASA. In the end of the design process, we will evaluate the result of the IA for VAL2 according with the EASA framework.,
- The user interface should be improved to ensure user acceptance and usability.
- 2 different scenarios have been tested in VAL1 to check the value of the IA according to task complexity. Pilots agree about the interest for the “diversion scenario”. The “rerouting scenario” is not very complex to be assisted by the IA.
- Proposed mock-ups for VAL1 have shown some failures and lack of functionalities. For VAL2, the representativity of IA should be improved to ensure:

  - translation between intentions and technical parameters;
  - the quality of the route assessment by the assistant, including best considerations of the available information for route evaluation;
  - The cognitive alignment, including communication/justification of proposals at adequate levels of abstraction and further interactions to support situation framing and decision-making.
- Integrate new information (for transparency and explainability)
- Pilots have appreciated the possibility to analyse and compare different solutions proposed by the IA. Multiple solutions should be presented with a set of immediate information.
- The way to present complementary information should be improved.
- Some "**training**" material should be provided informing which information is used to assess each intention and the safety of the solutions. Some examples could also be provided to enhance predictability/cognitive alignment previously. A small video should be provided before-hand illustrating the interaction with the assistant.
- For Safety, HF, Security and Legal evaluations of the concept, consider potential effects of the other functions and the interactions they imply.
- Conduct new interviews and workshops with pilots to determine the most concise and comprehensible way of presenting additional information on the interface. The aim is to avoid information overload and enable rapid analysis and understanding to facilitate decision-making. During this workshop, pilots will be asked to rank information based on its importance, clarifying which data should be displayed directly and which can be accessed on-demand, thereby enhancing comprehension.
  - The added value of customization of options presentation order, supporting different decision-making methods, may be investigated.

4.6 UC#2 Validation 1 Conclusion

Upon completion of VAL1, these are the main lessons learned that point to new design directions:

Table 10. UC#2 lessons learned

Insight	**Functional Requirements**	**Proposed solution for next iteration**
IA training and development	The representativity of IA should be improved.	Ensure translation between intentions and technical parameters, the quality of the route assessment by the assistant and the cognitive alignment.
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	The possibility of seeing whether a single solution	Multiple solutions should be presented with a set of immediate information.

and compare different solutions proposed by the IA	proposed with an AI trained by the pilots is enough to align intentions and propose other solutions to compare with demand.	
Various additional information was requested by the pilots	Analyse how to provide a summary of the additional information to not overload the pilot and ensure decision making quality and efficiency.	Investigate added value of customization of options presentation order, supporting different decision-making methods

4.7 UC#2 Appendix

CSUQ questionnaire in Peac²h

Peac²h https://app.peac2h.io/surveys/271/test_protocol

Mode Aperçu | Mode Test

1. Overall, i am satisfied with how easy it is to use this system.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

2. It is simple to use this system.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

3. I am able to complete my work quickly using this system.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

4. I feel comfortable using this system.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

5. It was easy to learn how to use this system.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

6. I believe i became productive quickly using this system.

1 sur 3 26/01/2024, 11:36

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

7. The system gives error messages that clearly tell me how to fix problems.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

8. Whenever i make a mistake using the system, i recover easily and quickly.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

9. The information provided (such as online help, on-screen messages and other documentation) with this system is clear.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

10. It is easy to find the information i needed.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

11. The information provided with the system is effective in helping me complete my work.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

12. The organization of the information on the system's screens is clear.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

13. The interface of the system is pleasant.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

14. I like using the interface of the system.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

15. This system has all the functions and capabilities i expect it to have.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

16. Overall, i am satisfied with the system.

Strongly agree ○ ○ ○ ○ ○ ○ ○ Strongly Disagree

Complete

Trust questionnaire

Peac²h https://app.peac2h.io/surveys/2632/test_protocol

Mode Aperçu Mode Test

For the below listed items, please read each statment carefully. Using the 5 point scale ranging from 1 (Strongly disagree) to 5 (strongly agree), select the answer that most accurately describes your feelings.

1. Generally, i trusted automated agents

Strongly disagree **Strongly agree**

2. Automated agents help me solve many problems

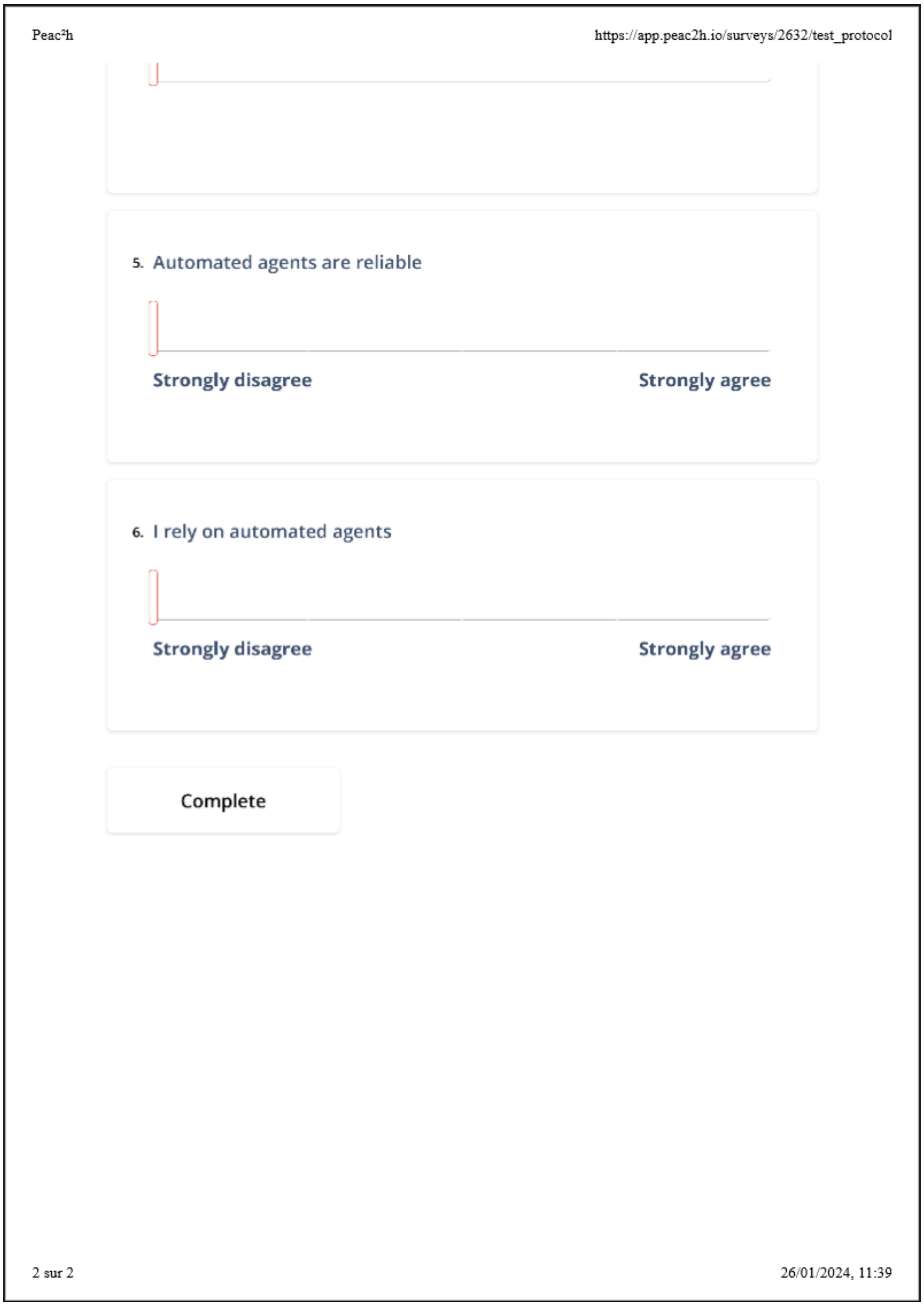
Strongly disagree **Strongly agree**

3. I think it's a good idea to rely on automated agents for help

Strongly disagree **Strongly agree**

4. I don't trust the information i get from automated agents. (R)

1 sur 2 26/01/2024, 11:39


Peac²h https://app.peac2h.io/surveys/2632/test_protocol

5. Automated agents are reliable

Strongly disagree Strongly agree

6. I rely on automated agents

Strongly disagree Strongly agree

Complete

2 sur 2 26/01/2024, 11:39

Open questions for the level of assistance

Peac²h https://app.peac2h.io/surveys/2633/test_protocol

Mode Aperçu | Mode Test

1. What do you think of 1B assistance? What did you like about it?

2. What did you dislike about the 1B assitance and how can it be improved?

Complete

1 sur 1 26/01/2024, 11:42

5. Use Case #3 – Urban Air Mobility

5.1 UC#3 Validation 1 Plan Summary

The primary goal of VAL1 is to investigate the operational concept of the intelligent assistant, the **Digital Assistant for UAM Coordinator (DUC)**, within specific scenarios. Additionally, VAL1 aims to examine the anticipated responsibilities of the UAM Coordinator, who is expected to engage with the DUC, and the operational concept of Urban Air Mobility (UAM) as applied in Stockholm, Sweden. A key focus is on understanding the interactions between the DUC and the **UAM Coordinator**, particularly in the context of human-AI collaboration to address the challenges presented in the VAL1 scenario.

Table 11. UC#3 performance targets

KPA	Category	KPI
System Performance	MoP	Task completion time, task accuracy, system availability, timeliness, synchronisation, data exchange accuracy
Human Performance	MoP	Use Experience and Usability: Workload, situational awareness, operational method, acceptability, trust
Safety	MoE	Adherence to safety procedures, lateral/vertical separation, speed restrictions, airspace restrictions

The goal of VAL1 comprises several sub-objectives centred around the three concepts of operations: 1) UATM in Stockholm, Sweden, 2) the DUC, and 3) the UAM Coordinator human role. VAL1 is the first human-in-the-loop simulation in the UAM use case that involves 8 air traffic controllers (ATCOs), who represent a potential group of end-users' representative for the role of the UAM Coordinator. So far, the focus has been on researching and developing the concept of operations (ConOps) for these three fundamental aspects in the UAM use case. As an extension of this foundational work, we have started a design process to implement these operational concepts in a scenario and simulator platform. VAL1 represents an initial milestone in this design work that will inform further design.

More specifically, VAL1 is intended to focus on:

- Communication and interaction between DUC and UAM Coordinator
- Human understanding of DUC

- Definition of the UAM Coordinator role, in terms of human role in the UATM scenario

VAL1 addresses a crucial research and development requirement by defining an operational concept that integrates human operator tasks with intelligent assistance to make key UAM operations feasible in terms of airspace and traffic coordination. This involves utilising scenarios to depict essential airspace operations and employing a prototype to symbolise a forthcoming IA in action. The primary focus of our R&D efforts is on gathering feedback on our (three) concepts of operation and interface design within the operational context of human and IA working together as airspace managers. Results are additionally expected to address which human-AI teaming strategy to target (EASA, 2023): cooperative or collaborative.

Our VAL1 strategy identified five R&D validation objectives (OBJ) and associated success criteria (CTR, see D6.1):

1. OBJ: To assess the operational feasibility and acceptability of the concepts (UAM, DUC, UAM Coordinator).
    a. CTR1: Positive feedback from the UAM Coordinator with respect to related tasks through cognitive walkthrough analysis. Acceptable quality of service, safety and workload.
    b. CTR2: The DUC can manage the traffic inside the related airspace in an acceptable way.
    c. CTR3: The HMI interface is user-friendly and provides the UAM Coordinator with necessary information.
2. OBJ: To assess the tasks and operating methods of the UAM Coordinator.
    a. CTR4: The content of the operating methods has been determined to be clear and consistent by UAM Coordinator and experts.
3. OBJ: To assess the UAM Coordinator timeliness of actions, workload, situational awareness, trust and acceptability.
    a. CTR5: The UAM Coordinator can perform tasks in an accurate, efficient and timely manner.
    b. CTR6: The UAM Coordinator workload is at an acceptable level.
    c. CTR7: The UAM Coordinator can maintain an acceptable level of situational awareness.
    d. CTR8: The UAM Coordinator has an acceptable level of trust and acceptability about the concept.
4. OBJ: To assess the effectiveness of the DUC in supporting the UAM Coordinator when a flight needs to deviate from its original route.
    a. CTR9: The DUC should provide enough support to the UAM Coordinator to handle the deviation/emergency without compromising safety.
    b. CTR10: The DUC should be able to provide information about the most suitable route, taking into account traffic, ground activity/availability, distance and airspace restrictions.
5. OBJ: To assess the DUC HMI interface and information requirements.
    a. CTR11: Positive feedback from the UAM Coordinator.
    b. CTR12: Positive feedback from the UAM Coordinator on the information provision.

To explore the research objectives, two scenarios in Stockholm, Sweden, were scripted.

- Stockholm Scenario 1: Re-routing of an air taxi due to a medical emergency. Single passenger air taxi transporting a passenger from location Globen that becomes unresponsive and is rerouted to a suitable location where medical treatment is available (Karolinska University Hospital).
- Stockholm Scenario 2: Re-routing of a package delivery through Bromma city airport ATM Control Zone. Delivery drone transporting a package from location Västberga to location Hässelby Strand and requests a re-route during the flight.

VAL1 utilises the UTM City software program (Lundberg et al., 2021) as it offers a controlled and cost-effective approach to validate the DUC concept and prototype. It enables iterative testing, refinement, and optimization of functionalities and performance within a simulated urban airspace. Gathering feedback from expert air traffic controllers (ATCOs) working as an UAM coordinator on the UTM City platform provides valuable insights, contributing to the ongoing development of the three concepts.

The European Union Aviation Safety Agency (EASA) provides guidance on Human-AI Teaming (HAT) with six categories describing Human-AI partnerships. The DUC IA in UC3 aims at providing support for the following category:

- 2b – Collaborative agent - an autonomous agent that works with human colleagues, but which can take initiative and execute tasks, as well as being capable of negotiating with its human counterparts

5.2 Deviation from Initial Validation Plan

A change was made to the scenario design in VAL1. Instead of exploring two scenarios in Stockholm separately, we decided to create a single scenario that contained both Scenario 1 and Scenario 2. To avoid confusion, we hereafter refer to Scenario 1 as Event 1 and Scenario 2 as Event 2. The decision to use a single scenario that contained two events (i.e., previous Scenario 1 and Scenario 2), rather than two separate scenarios, was taken to reduce the effort in scenario development and simulator implementation. While the core events required relatively little effort to implement in UTM City, the surrounding "noise" traffic and activities required considerable effort.

No changes were made to stipulated OBJ (research objectives) and CTR (success criteria) in D6.1 (and reflected in the previous section). However, two additional overarching research questions were added for exploring 1) participants' responses on interactive storytelling as an approach for explaining how DUC works and 2) participants' perceptions of DUC's human-AI teaming characteristics.

With respect to storytelling, the overarching research question was: How can the interactive storytelling elements be optimally (for target end user) designed to effectively convey the backend processes (how the system works - how the suggestions are derived) of AI intelligent assistants, increasing user's understanding and trust in the intelligent agent?

In particular, we were interested in learning:

- How do participants perceive the level of detail provided (theory of CLT by Trope & Liberman (2010) in the storytelling elements?
- How does the users' trust and understanding vary when interacting with the storytelling elements, as compared to without storytelling elements?

With respect to human-AI teaming, the overarching research question was: How can the DUC be designed to be perceived as a team member by air traffic controllers acting as an UAM Coordinator in UATM?

In particular, we were interested in learning:

- What are desirable teaming characteristics of DUC in working with the UAM Coordinator to handle UAM vehicle rerouting situations?

- What similarities and differences are there between the ATCO role and UAM Coordinator role?

5.3 Concept development

The UAM Use Case proposes a UAM Coordinator working position as envisioned in 2030. The work conducted in defining the HAIKU UAM ConOps is found in the UC#3 Appendix. The following text is a portion of the content provided in the UC#3 Appendix. At the time of this report, there is no current UAM or UATM in operation. Therefore, the first step was to determine a concept of operations for UAM. The developed UAM concept was based on that of the U-space concept of operations proposed by the European research project: Concept of Operation for EuRopean UTM Systems - Extension for Urban Air Mobility (SESAR JU, 2022). Urban Air Mobility (UAM) can be defined as the safe and efficient air traffic operations in a metropolitan area for manned and unmanned aircraft systems. As a starting point, there is a need for defining the UAM ConOps in which the use case and scenarios are to be realised. The HAIKU UAM use case has explored several existing UAM ConOps.

The HAIKU UAM ConOps is based on the CORUS-XUAM ConOps. CORUS is a Horizon 2020 project funded by the SESAR Joint Undertaking that has produced four iterations of U-space ConOps, starting in 2017. The fourth iteration focuses on user needs of UAM. According to CORUS-XUAM, the term UAM is defined as air operations above urban areas (can be part of the flight) in U-space airspace where tactical separation is needed to cope with high traffic density to ensure separation. Traffic is expected to be mixed, including aircraft with limited range and incapable of flying IFR or VFR. UAM ConOps and research on UAM or similar, generally assume that the UAM system will be fully automated. Detection and avoidance (DAA) are envisioned to be automated: vehicles will have their own sense and avoid solutions. There has been very limited focus on exploring the human role in envisioned UAM landscapes. UTM is envisioned to be autonomous involving as least human as possible. If anything, the human is envisioned to oversee the operation and be able to assist, when necessary (supervising role), or the human could be a fallback system when automation fails. HAIKU sees a need for humans engaging with the UAM system to accommodate the link between ground and airborne activities.

In HAIKU we envision a need for human involvement in safely and efficiently managing the airspace above cities and coordinating UAM with ground activities. We envision this human role as a UAM Coordinator.

Role and key competencies for the UAM Coordinator

The role and responsibility of the UAM Coordinator would be to provide U-Space services to UAS and UAM operators, including geo-awareness services (e.g. configuration of airspace constraints and temporary restrictions), flight authorization (managing and authorising flight plans), and traffic monitoring. The key purpose of the UAM Coordinator is, as the name suggests, to coordinate between stakeholders involved in the UAM ecosystem. This is in contrast to ATCOs who manage traffic.

The UAM Coordinator would probably be an actor employed by the local USSP. In HAIKU we envision that each city would have its own local USSP, where a single or multiple UAM coordinators could be responsible for the airspace overhead the city (e.g., if there are multiple USSPs within a city). The UAM coordinator has some resemblance to ATCOs in that both provide a service to aircraft. A difference, however, is that aircraft in the U-space, in contrast to controlled airspace in ATM, are envisioned to operate autonomously to solve route monitoring and CD&R without the need for human intervention. As such, a more suitable comparison might be with road traffic management operators. In road traffic management, cars operate 'autonomously'. Instead, the road traffic management operators focus on flow management, managing constraints (e.g., closing roads), and reacting to emergency situations. Similarly, the role of the UAM Coordinator is envisioned as a requirement for managing the flow of aircraft, geo-fence provisions, reacting to emergency situations, and coordinating dependencies between ground-based activities and aircraft operations.

Note that many of the USSP services are assumed to be highly automated, but some are likely to require human involvement to a higher degree due to the uncertain and dynamic nature of the service. One such critical service focused on in HAIKU is emergency services.

We envision the need for a UAM Coordinator role to be especially important in the growth and development of UAM and U-space services (i.e., as U-space operations transitions U2 - initial services, to U3 - advanced service, toward U4 - full services). As such, the target for VAL1 activities is 2030. However, as UAM approaches full services (U4), the need for a UAM Coordinator role may diminish with increased autonomy.

UC#3 Demonstrator and Assistant Features

For VAL1, a proof-of-concept UATM working position for the UAM Coordinator has been developed. In addition, a proof-of-concept prototype of the DUC has been developed and integrated with the HMI. As a demonstrator for the UAM use case a video recording is provided that captures the medical emergency situation where a sick passenger in an unpiloted air taxi is rerouted to a hospital.

The DUC prototype is being developed having capacity to exchange traffic data and e-VTOL and drone flight plans between the UTM City and the SOMA-AI platform. Through the DUC dialogue window, the UAM Coordinator can ask for explanations for recommended actions. In the future, the UAM Coordinator could possibly allocate tasks and duties to DUC, adjust strategic performance parameters for DUC (e.g., prioritisation rules, separation criteria) and change the level of automation of DUC.

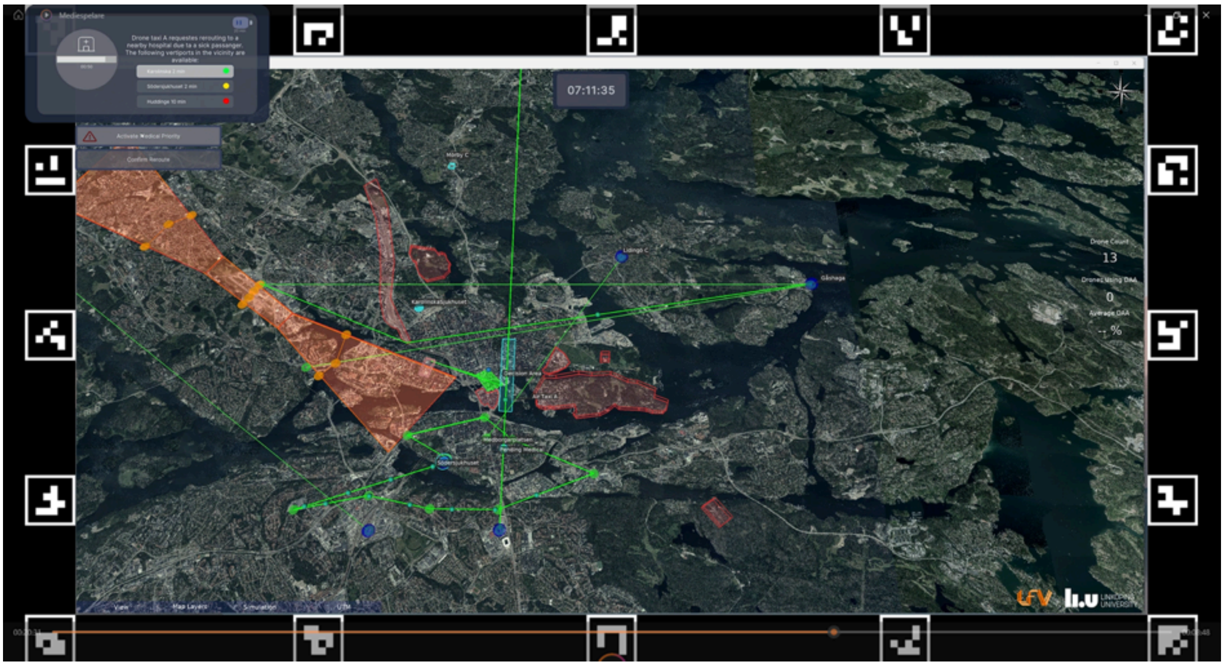


Figure 31. UC3 DUC dialogue window with suggestion on how to solve the medical diversion (upper left corner)

In Figure 31, the DUC dialogue window is shown. The dialogue window in the upper left corner presents three hospitals in the vicinity that can accept the Air taxi and patient. The location of the hospitals is indicated by cyan filled circles on the map. DUC provides an estimated flight time to each hospital in minutes from the current position of the Air taxi. The green, yellow, and red dots next to the hospital name (dialogue window) indicates the most suitable hospital (according to DUC), taken into account parameters such as staffing/specialist and waiting time, vertiport conditions (stands, battery charging etc), and flying time (green = best; red = worst). The validity of the recommendation, and dialogue window, is indicated by the round time bar with the hospital symbol. This time corresponds to the transparent cyan rectangle symbol overlaid on the Air taxi's route (named 'Decision area') on the map view. The Air taxi route is the green line originating at the blue dot on the lower part of the map and going almost straight up in the map (beyond the current zoomed in view). The current position of the Air taxi is along the route, south of the 'Decision area', indicated by a dot with a label 'Pending Medical'. The green lines on the map indicate flight plans. Blue dots indicate vertiports from where vehicles take off or land. Red zones are geofences, or restricted areas, that constrain or prohibit vehicles movement. The orange zone represents the Bromma city airport and the boundary to ATM. These key symbols are illustrated in Figure 32.

For purposes of explaining the UAM concept, DUC, and UAM Coordinator role, we used a Storytelling approach to create "stories" describing these. The Storytelling explainer consisted of scripted video recordings combining visual media and narration to explain a particular construct (e.g., DUC). The Storytelling explainer was provided on a separate display, next to the UTM City

HMI. The Storytelling explainer would provide explanations for the participant to view, timed to specific happenings occurring in the simulation.

The Storytelling explainer contained four elements.

- Storytelling element 1: provide the UAM Coordinator with the overall understanding of the UAM concept. The story covered the basic UAM terminology illustration: point-to-point, geofence, vertiport. For each term audio, animation and map explanation was provided. The length was about 3 min.
- Storytelling element 2: explain to the UAM Coordinator what the DUC is and how responsibilities are divided between DUC and UAM Coordinator. The story covered an introduction of DUC, its abilities and limitations. Additionally, it covered the responsibility of the UAM Coordinator and DUC. The length was about 2 min.
- Storytelling element 3: introduce Event 1: 'Re-routing of an air taxi due to medical emergency' by briefly explaining the technical specifications of air taxis i.e., sensors that monitor passengers' health parameters. The length was about 1 min.
- Storytelling element 4: explain to the UAM Coordinator the backend and decision-making processes of the DUC and how it derived the recommendation communicated in the pop-up message on the simulator screen in relation to the Event 1. It included illustrations of different actors of the scenario (air taxi operator, emergency response centre and DUC) and their system backend process. The length was about 3 min.

Figure 32. Storytelling explainer in UC3

5.4 UC#3 TRL

In UC3, there are four system components related to the intelligent assistant that are explored. They are reported below, including their initial TRL:

- AI underlying the UAM Intelligent assistant DUC, TRL 1,

- HMI - Interactive touch-based interface and traffic visualisation tool (UTM City) and IA interface, TRL 3,
- System and convolutional neural networks for image recognition to monitor and process object movement (SOMA-AI) and overlay functionality, TRL 3,
- XAI UI – Storytelling explainer for explaining how the intelligent assistant works, TRL 1.

In VAL1, our experimental system has utilised two previously developed components (from other projects): the UTM City HMI for visualising UAS/UAM traffic on a map display, and the SOMA-AI for providing overlays on the UTM City HMI. Note that the image recognition and object movement functionalities of SOMA-AI were not implemented in VAL1. Each of these components has advanced in their TRL throughout earlier projects, becoming more effective tools in their respective contexts. As an example, UTM City has been used as a tool for exploration and design (TRL8/9), but its use as airspace control software has only been demonstrated (TRL5/6).

The IA interface, which is embedded in the UTM City interface, is novel work. This part, and thus also the full HMI, should be considered TRL3 Experimental proof of concept as it demonstratively works. The goal is to also validate that it works, raising it to TRL4 Technology validated in the lab. The overlay functionality is similarly considered TRL3 Experimental proof of concept as it demonstratively works, but should also be tested with the image recognition and object movement functionalities.

The AI underlying the DUC (i.e., backend of the DUC) is only conceptually considered at a very general level in VAL1, related to high-level human-AI teaming requirements. For VAL2, we will seek to develop a more detailed technology concept building on basic principles from theoretical frameworks.

The Storytelling explainer (XAI UI) concept is a novel attempt to combine storytelling and construal level theory to provide explanations for the underlying decision making of an intelligent assistant. The implementation is simple as it makes use of scripted video recordings combining visual media and narration, and therefore currently only reaches TRL1.

5.5 UC#3 Validation 1 Methods

Between the period of D6.1 submission and this deliverable, our focus has been on preparing VAL1. Core activities include iteration and refinement of 1) the UAM concept of operations, the DUC concept of operations, and the UAM Coordinator concept of operations. Combined these make up the three key components of the UAM Coordinator working position. During the fall of 2023 we have:

- Finalised a scenario in Stockholm U-Space with two key events (Event 1 and 2).
- Continued development of the HMI (i.e., UTM City) to exchange traffic data and drone flight plans with the SOMA-AI platform.
- Designed and implemented dialogue windows and synthetic voice alerts in the HMI as DUCs primary means of communication with the UAM Coordinator.

- Designed and implemented Storytelling elements with the HMI for the purpose of introducing and explaining the UAM concept and how DUC works.

Moreover, we have conducted several internal workshops and technical tests in preparation for VAL1. Test runs of VAL1 were conducted on January 25th and 29th. During this test we tested scenarios, validation procedures, debriefing interviews and questionnaires.

Validation design

VAL1 was an explorative validation study incorporating both qualitative and quantitative elements. As such, our experiment focused towards exploring and understanding different concepts as opposed to decisively confirming or establishing the characteristics, capabilities, or outcomes of a particular concept. Explorative approaches are common in early stages of development when there are uncertainties about the capabilities, functionalities, or performance of a system. This is the case of Urban Air Mobility which is in the making. At this stage, there is a need to gain insights, gather information, or uncover potential issues in a relatively open-ended manner. This requires a more flexible and investigative approach, where the primary goal is to discover and learn rather than to confirm or verify predefined expectations.

The validation study was a between group design with the Storytelling explanation as an independent variable. The simulation consisted of a single scenario that each participant encountered twice. For Group A, the Storytelling explanation was provided together with the first run of the scenario. In the second run of the scenario, no Storytelling explanation was provided. For Group B, the order was the opposite.

Measurements were collected in terms of observations and simulation measures, three different questionnaires, and two debriefing interview guides. Audio, screen capture, video, eye-tracking, and image recording devices were used to capture participants' communication and behaviour.

Participants

Eight experienced ATCOs provided by LFV Sweden participated in VAL1. Age ranged from 24 to 57 years (M = 42.6). All ATCOs were working at Malmö ATCC. Experience ranged from 0.4 to 34 years (M = 16.9). While all participants had a basic understanding of UAM in general (courses provided internally within LFV), they had no previous knowledge or experience of HAIKU, UAM, DUC, or the UAM Coordinator. Participants in Group A (Storytelling condition first) varied in age from 38-47 (M = 42.8) and experience from 2.1 to 25 (M = 16.4). Participants in Group B (Storytelling condition last) varied in age from 24-57 (M = 42.5) and experience from 0.4 to 34 (M = 17.5).
The choice of ATCOs as participants and representatives for the UAM Coordinator in VAL1 was not obvious. In relation to developing the UAM Coordinator concept and working environment, and discussing roles and responsibilities, it was clear that the UAM Coordinator role is uniquely different from that of current day ATCOs. Differences include low level urban airspace operations, manned and unmanned e-VTOL and drone operations with very different vehicle performance characteristics and behaviours compared to aircraft in ATM operations, and increased coordination with ground activities. However, it was also clear that the UAM Coordinator role has many similarities with ATCOs and overlapping high-level tasks such as traffic monitoring and traffic flow management, ground-air coordination, and handling emergency situations. ATCOs were considered the most suitable actors representing knowledge

and skills considered important for the UAM Coordinator role, available in current day operations.
ATCOs were chosen as participants because:

- ATCOs understand aerodynamics and traffic movement in the air.
- ATCOs are envisioned as the candidate population to assume the role as UAM Coordinator.
- ATCOs are experienced in coordinating with other sectors, stakeholders. For example, in case of a flight diversion from one airport to another. Such routines and processes could be investigated during the validation use cases. Who do they contact first, what happens after this in the chain, what preparations are made at the new and previous airport.

However, future work should consider other actors when exploring the roles and responsibilities of the UAM Coordinator. In particular:

- Tower ATCOs, particularly Bromma city airport, would have useful local procedures knowledge.
- Road traffic monitoring operators would know how to handle deviations on a strategic level, with blocked roads etc.
- Joint Rescue Coordination Centre (JRCC) and SOS Alarm operators would be able to provide information of these steps, who are first medical experts in case of e.g., cardiac arrest.

Simulator

The UAM Coordinator working position consists of a situation display covering the Stockholm U-Space airspace and relevant UAM components (i.e., vertiports, geofences, and UAM vehicles). The UAM Coordinator uses a mouse to orientate and move around in the Stockholm map. The DUC is integrated with the situation display using overlay software. The UAM Coordinator has a communication system (i.e., telephone) for communicating with other stakeholders such as ATM (e.g., Bromma city airport), JRCC, drone operators etc.

The UAM Coordinator working position was created using a portable simulator platform. The simulation platform consisted of

- UAM HMI (UTM City): a portable computer connected to a 4K external screen.
- Storytelling explainer: a portable computer connected to a 4K external screen positioned next to the HMI.
- A fast dial telephone for the UAM Coordinator working position.
- A 'rest of the world' position with a fast dial telephone (for experimenters).
- Tobii 3 portable glasses eye tracking system.
- Screen recording and overlay kit consisting of a laptop with external capture card.

The UATM HMI is using UTM City as an interactive visualisation of the U-space in Stockholm and simulates UAM vehicle traffic/services and airspace restrictions/rules on a map with a dashboard. The Joint Cognitive Framework (JCF) Editor, developed at LIU, is used as a combined scenario planning and generation platform that interacts with the UATM HMI, and an analysis tool of human-automation/AI interactions. The overlay software, a component of SOMA-AI created at LIU, displays information from the DUC and gets input from the operator by adding overlay elements to the situation display. Overlays are triggered from JCF-E and inputs are forwarded back to the editor.

The Storytelling explainer was built in the software for an interactive projection and editing, WATCHOUT Dataton. The storytelling explainer contains four elements that were developed by using different visual formants (motion design, filmmaking, stock footage use, AI voice generation, etc.). Every element was located on the separated timeline and connected to The JCF-E with a trigger on the timeline, so it could be triggered in the specific order in the scenario. The storytelling part was built only for Event 1: Re-routing of an air taxi due to a medical emergency. Every time when the event with a storytelling explainer starts it immediately triggers the first element. After some pause it triggers the second element. When the first air taxi starts from Lidingö it triggers the third element. Finally, three minutes before the first pop-up with choices it triggers the fourth element. Attention guidance between two screens was designed with storytelling black screen lighting up to bright white when the story was triggered.

Intelligent Assistant

The DUC IA aids the UAM Coordinator in overseeing U-space operations. It handles most routine tasks, such as flight authorization, traffic 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.
- Informing and communicating with other stakeholders, including other traffic and emergency response teams on the ground and in the air.
- Following emergency procedures, such as adjusting dynamic safety boundaries, changing flight prioritisation, and coordinating emergency routes.
- Establishing priority criteria dynamically for different types of flights.
- checks for anomalies

In VAL1, DUC was realised by a Wizard of Oz approach focusing on DUC actions in relation to Event 1 and Event 2. For this purpose, DUCs tasks and actions were scripted beforehand. A graphical user interface, using overlay software, for DUC communications was designed and implemented in the HMI as pop-up dialogue windows.

Scenario

A single scenario was created, containing two events: Event 1 and Event 2. The scenario takes place in the Stockholm U-space area, which was developed and implemented from scratch in the UTM City platform to create the UATM HMI.

For purposes of the first validation, a U-space was created for the Stockholm metropolitan area. This comprised creating and placing (in the Stockholm area) vertiports, ATM restriction areas (Bromma city airport), designated traffic zones (i.e., logistic hubs and sport arenas), obstacles (buildings/constructions), geo fences (restricted zones around sensitive locations of military, rescue services, or governmental importance), and additional areas of interest that pose constraint to certain types of operators. This airspace was then implemented in the UTM City HMI. The next step consisted of populating the airspace environment with traffic and events. Routes and services (i.e.,U-plans) were created for air taxi operators, surveillance drones, sightseeing operations, delivery drones, and medical transports. As part of this work, a storyline was developed as a basis for what became the scenario explored in VAL1. The scenario comprised two situations that we had determined would require coordination between the human actor (the UAM Coordinator) and the AI actor (the DUC). The two situations, referred to as Event 1 and Event 2, were scripted to occur in sequence.

The following assumptions were made for the scenario:

- U-plan is approved and strategically deconflicted.
- DUC is part of the USSP system but standalone.
- Separations between airspaces users in U-spaces are in place.
- Regulations supporting ATM/UAM integration.
- Tactical deconfliction and DAA is present.
- DUC has information about vertiports and other landing sites.
- DUC has information about hospitals and their capabilities and services.
- DUC has adequate weather information.
- Bromma CTR has not the same design as today, it is adapted to U-space airspace

Event 1 – Re-routing of an air taxi due medical emergency: Single passenger air taxi e-VTOL transport flying a direct route from south position Globen to north position Täby Centrum in Stockholm, Sweden. Near position Hammarbyslussen the air taxi detects autonomously through sensors that the passenger has become unconscious. Information about the medical situation of the passenger is sent to the air taxi operator. The situation requires an emergency deviation to a hospital. DUC is notified of the emergency and notifies the UAM Coordinator by highlighting the air taxi in distress and suggesting a change of flight priority. Simultaneously, DUC queries nearby hospitals and presents the UAM Coordinator with a list of alternative, and suitable, hospitals that can accommodate the air taxi and passenger. The alternatives are communicated to the UAM Coordinator in a dialogue window. The UAM Coordinator is given 1 minute to decide on which hospital to direct the air taxi to. Once a decision is made, DUC coordinates and implements the new route.

Event 2: – Re-routing of a package delivery: A logistic company situated in position Västberga sends a delivery drone with a package to position Hässelby strand in Stockholm, Sweden. At approximately position Södra Ängby, the drone operator sends a request to change destination for the package delivery to a new destination: Huvudsta. The DUC processes the request and calculates different routes to the new destination. The shortest routes require transitioning through Bromma city airport restricted airspace controlled by ATM. A route around Bromma city airport restricted airspace, however, would result in a considerably longer flight. DUC requires the UAM Coordinator to contact the Bromma city airport ATCO to get permission for the delivery drone to enter and cross the controlled airspace. Following the UAM Coordinator's decision on how to act, the DUC responds to the delivery drone operator's request and sends out a new route. Event 2 occurs after Event 1 has ended.

During the scenario, one external call to the UAM Coordinator from external actors was made. The Railway inspection manager called and informed of a "signal error" and the need for a drone inspection of the rail over location Tomteboda.

In addition to the key events, several background activities had been scripted to make the scenario more complex and realistic. This included opening of restriction zones, opening of vertiports, and a gradual increase in traffic.

Once the UAM concept had been sufficiently developed and implemented in UTM City, we gradually shifted focus to the envisioned human and AI actors working together to provide U-space services: the UAM Coordinator and DUC. Therefore we had to first sketch out the concepts of operation for these two actors, their roles and responsibilities, and how they are to communicate and interact. As a result, the UAM Coordinator working position was developed. The coordination between DUC and the UAM Coordinator was scripted as points of interaction between the two actors in a table of tasks. The Table 12 below is an excerpt from the table of tasks.

Table 12. Table of tasks for Event 1 and Event 2 that specifies interactions between DUC, UAM Coordinator, and a third sector (e.g., UAM vehicle operator).

EVENT 1		
DUC	**UAM Coordinator**	**To/from external system (other than DUC or UAM Coordinator)**
Scans for anomalies		
		Emergency is sent from the taxi drones passenger sensors
Detects an emergency on a taxi drone		
Notifies UAM Coordinator		
	Receives a heads up (warning)	
	Acknowledges the message	
Sends additional information		
	Turns out to be a sick passenger (heart problems)	
	Verifies and approves the emergency	

	The verification implies that other drones in the area may receive a recalculated route	
This information is forwarded to USSP and CISP system		Receives information from the DUC
	Makes a decision – the drone continues on its precalculated (and deconflicted route)	
	Initiates check of hospitals that can accept this drone with a sick passenger	
Calculates which ERs that are in the vicinity of the taxi drone's route		
Sends out a request to those ERs		Request sent to the hospitals
		Answers, with possible restrictions
Calculates best option based on answers from the different ERs		
Presents the 2-3 best options with potential restrictions		
	Makes a decision based on the different options and UAM Coordinators experience	
	Karolinska Solna is considered the best option	
Receives the message about decision		
Sends messages to Karolinska Solna		
		Message received and booking confirmed
Confirmation received		
	Initiates a re-calculation of the route	
	Also approves a request of “emergency route”	
Calculates and sends (to USSP) the requested emergency route		
		New route received. Re-calculates this and other flights since this taxi drone has priority.

		Points out that this new “emergency route” will enter No fly zones. And permission is required
Receives this information		
Notifies UAM Coordinator of new route and No fly zones		
	Approves request to enter No fly zones	
Approval is received		
New route, approval and new landing site is sent to air taxi operator		
		Drone receives new U-plan
		Drone applies this new route
ETA is forwarded to Karolinska Solna.		
		ETA for Karolinska Solna ER is received
Displays that all the steps are done. And that the drone has accepted the new emergency route		
	Monitors the flight	
		Drone has landed and the passenger is unloaded. Confirmation sent.
Receives a confirmation on the landing and taking care of passenger at Karolinska Solna		
Displays this confirmation		
	Ends the monitoring of the flight	
Event 2		
DUC	**UAM Coordinator**	**To/from external system (other than DUC or UAM Coordinator)**
		The package delivery company has a request, re-routing of an already airborne drone
	Receives the request	
	Acknowledges the request and sends it to DUC	
Starts calculation of alternative routes.		

Calculates with a lot of parameters		
Sends a request about battery capacity		
		Company receives the request and answers it
Calculates the shortest, the fastest and the most battery saving routes.		
Also takes adequate weather into account		
Checks alternate vertiports and emergency landing sites		
Presents these calculations		
	Displays information that alternative new routes are sent to package delivery company	
		Two options are presented.
		Option 1: Shortest way but involves waiting time to cross thru SB approach (geo zone with grid)
		Option 2: Longer way, low flying over the water. No waiting times
		Decisions about preferred alternative
		Sends back selected alternative
	Informed about chosen alternative	
Receives selected alternative (1)		
Recalculates again taking into consideration the calculated position where it's going to switch to the new U-plan		
Parameters such as vertiports, battery etc is calculated again		
Sends new U-plan to USSP/CISP		
At the same time calculates the ETA at geo zone border		
	Heads up that a request soon is coming about crossing Stockholm/Bromma airport (ESSB) geo fence zone	

		New U-plan is tactical de-conflicted by USSP/CISP
		ETA at geo zone is received and perhaps adjusted according to the deconflicted U-plan
		Amendments are done and sent back including the amended ETA
Receives the amended U-plan		
Requests to cross SB geo zone is sent		
	Approves the request to be sent to ATM system	
		Request received by ATM and holding point and available crossing time is sent back
Receives the adjusted times and send this to UAS		
		UAS changes route according to approved U-plan
	Displays information about re routing	
	Informed about chosen alternative	
Receives selected alternative (2)		
Recalculates again taking into consideration the calculated position where it's going to assume the new U-plan		
Parameters such as vertiports, battery etc is calculated again		
Sends new U-plan to USSP/CISP		
		New U-plan is tactical de-conflicted
		Amendments are done and sent back
Receives the amended U-plan		
Sends this plan to package delivery company		
		UAS changes route according to approved U-plan
	Displays information about changed routing	

With the table of tasks as a reference, the means for communicating and interacting was designed and implemented in the UTM City HMI. DUC communicates with the UAM Coordinator through dialogues windows in the HMI and voice. The UAM Coordinator communicates and interacts with the DUC through dialogue windows in the HMI.

Measures

The following data collection methods were used:

- Think-aloud
- Observations
- Questionnaires
- Debriefings
- Eye-tracking data
- Audio recording
- Video recording
- Screen capture recording

Three questionnaires were used: Demographics questionnaire; Storytelling questionnaire; and DUC questionnaire. The demographics questionnaire asked participants about their age, experience, and ratings held. The Storytelling questionnaire consisted of eight items:

1. To what extent do you think that you collaborated with the DUC to solve the situations in the simulation? (1-5 scale)
2. To what extent do you feel the explaining stories contributed to this interaction? (1-5 scale)
3. The explaining stories made DUC (where 3 is neutral) (1-5 scale)
4. The explaining stories made my trust in DUC (where 3 is neutral) (1-5 scale)
5. DUC managing the sick passenger re-routing made me (where 3 is neutral) (1-5 scale)
6. My degree of control of the U-Space traffic made me feel (where 3 is neutral) (1-5 scale)
7. I feel that the explaining stories were useful/useless (1-5 scale with Story 1-4 rated separately)
8. Mark on a scale of 1 to 5, how would you rate your overall experience with the explaining stories? (List of experiences: Unfriendly, Boring, Confusing, Not interesting, Unpredictable, Complicated, Slow, Demotivating, Does not meet expectations)

The DUC questionnaire consisted of fourteen items and Van der Laan et al.'s acceptance scale (Van Der Laan et al., 1997):

1. I accepted DUC's recommendation in the sick passenger rerouting situation without inspecting the situation (1-7 scale)
2. DUC was very helpful in handling the sick passenger rerouting situation.
3. I would like to know how DUC reasoned to derive the recommendation in the sick passenger rerouting situation.
4. I accepted DUC's recommendation in the delivery drone rerouting situation without inspecting the situation.
5. DUC was very helpful in handling the delivery drone rerouting situation.
6. I would like to know how DUC reasoned to derive the recommendation in the delivery drone rerouting situation.
7. I felt comfortable working with the DUC.
8. I would like to have more control over DUC.
9. A system like DUC would make my job less rewarding.
10. A system like DUC would allow me to focus on important aspects.
11. How similar is the role of the UAM Coordinator to that of an Air Traffic Controller?
12. DUC lowered my workload.
13. It felt like DUC and I worked together as a team, where DUC was my team member.
14. I felt that DUC was a support system or tool that I could use.
15. Acceptance scale: Please tick a box on every line (9 lines with a 1-5 scale).
    I find the DUC… (useful/useless; pleasant/unpleasant; bad/good; nice/annoying; effective/superfluous; irritating/likeable; assisting/worthless; undesirable/desirable; raising alertness/sleep-inducing)

Debriefing protocols were developed separately to assess teaming aspects and Storytelling explainer aspects. The teaming debriefing consisted of eight items:

1. Was it clear which specific tasks DUC was doing? Why?
2. Are there certain things you think you should have done and not the DUC? Why?
3. Describe how DUC communicated?
4. How can it (the communication) be improved?
    a. What did you think of the method/way that DUC communicated, through window prompts with glyphs/graphical elements/symbols?
    b. Could you think of other ways for communication?
5. Was there any information or updates you would have wanted more?
6. Which relation would you think is most suitable for the Human DUC interaction?
    a. Like an assistant? (1-5 scale)
    b. Like colleagues? (1-5 scale)
    c. Like a leader? (1-5 scale)
    d. Other?
7. What do you think is most important for teamwork in this context?
    a. Closed-loop-communication
    b. Shared mental models
    c. Mutual performance monitoring
    d. Implicit communication
    e. Adaptability
    f. Backup behaviour
    g. Team leadership
    h. Boundary awareness/understanding of the system and its limits
8. Can you think of any example on how this concept can be further developed in the future?

The Storytelling explainer debriefing contained five sections covering 1) general questions on participants’ perspective and experience from watching the storytelling elements; 2) understanding of storytelling elements; 3) trust perceptions of DUC; 4) the level of detail to assess the storytelling element’s impact on cognitive load; and 5) suggestions on improvements.

Audio, video, eye-tracking, and screen capture recordings captured all aspects of participant’s communication and physical interactions in the UAM Coordinator working position. The recordings are mainly used to assist in qualitatively analysing the scenario and key events to understand what happened and what the participant did and did not do. Eye-tracking aimed to monitor how the operator collected information, as well as measure time, in terms of how quickly the UAM Coordinator noticed when the DUC displayed information, and the time difference between getting the information and when a decision was taken.

Procedure

Figure 33 b below illustrates the procedure used for VAL1. The entire simulation took 3 hours per participant. First, participants were introduced to the research team, read and signed a consent form (which had been sent out to participants via email prior to the simulation) and the demographics questionnaire. Secondly, the research team presented the HAIKU project, followed by an introductory presentation of U-space in general and UAM use case in Stockholm specifically. This also included a presentation of their role as the UAM Coordinator and the DUC.

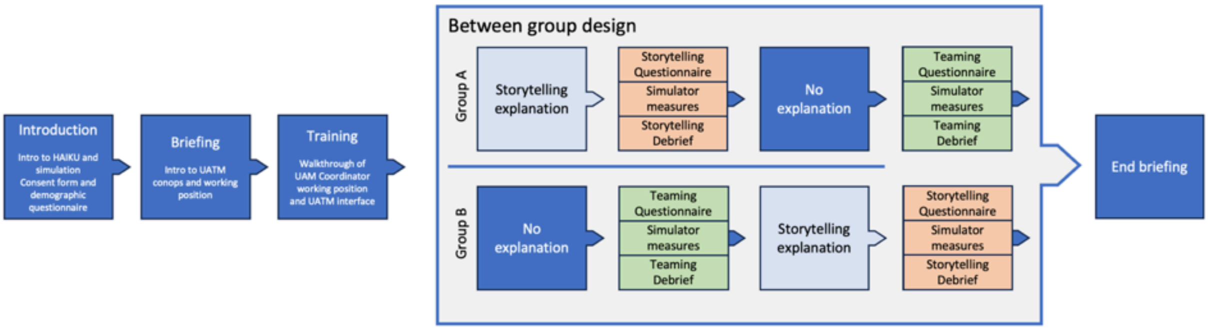


Figure 33. Overview of VAL1 procedures for UC3.

Next, participants received a simulation briefing consisting of:

- a description of the simulator and what was required from the participant,
- scenario specific reference material: Stockholm map, symbols legend, and special notices (i.e., expected events that can affect the U-space).
- Protocols for contacting other actors (using the telephone)

Participants were introduced to the simulation with the following premises:

"It is summer in Stockholm, Sweden, and you (the UAM Coordinator) start his/her working shift just before 07:00, which is the time the Bromma tower will open."

After the briefing, participants received a training walkthrough of the UAM Coordinator working position, briefing material such as the scheduled flights for the shift, a summary of important events such as the likelihood of inspections of roads, railways and masts, significant ground events such as preparations for a concert over Djurgården, a symbol key (Figure 34) for the operative concepts such as point-to-point services, point–loop-services, point-to-scan-area services, constraints (geofences and airport gates), vertiports, points of interests, and decision zones, the UAM HMI, and the Storytelling explainer HMI. Participants were instructed how to interact with the UAM HMI, how to zoom, move the map and align the chart to the north, and Storytelling Explainer. This part of the validation took around 50 minutes.

.

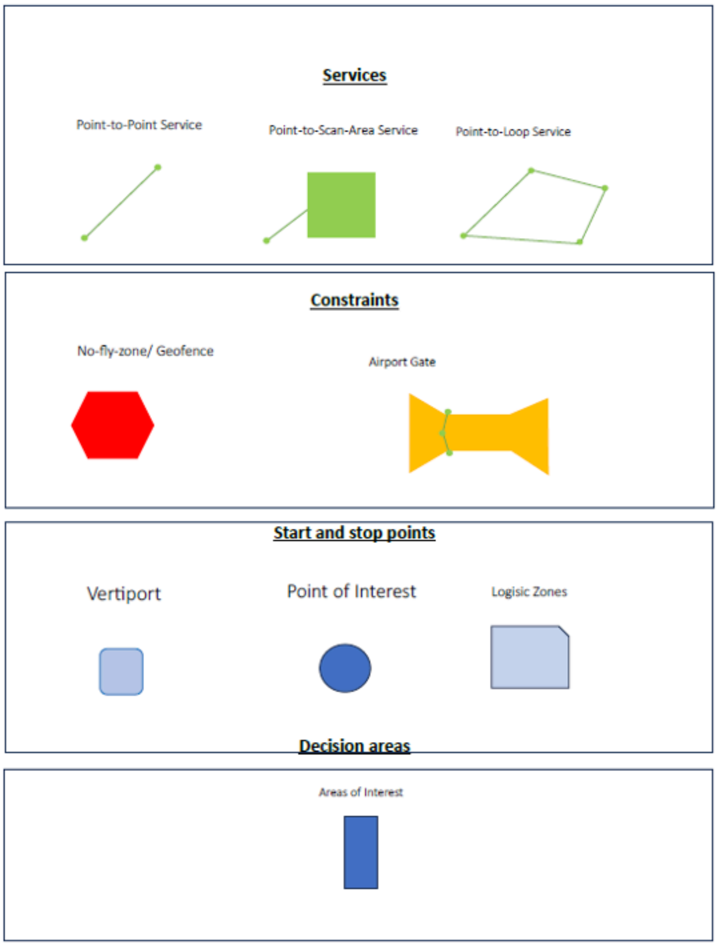


Figure 34. Symbol key with information about UAM related symbols in the UATM HMI

After a short break, the actual simulation started. Participants were divided in two groups that played the same scenario twice. Before each run, the eye-tracking was calibrated and tested. The scenario contained two scripted events (Event 1 and Event 2) and ran for about 30 minutes. We were particularly interested in how participants, with DUC, solved the two events in the scenario. The difference between scenarios was whether the Storytelling explainer was provided or not. Participants in Group A first encountered the 'Storytelling explainer' condition, followed by the 'no Storytelling explainer' (no explanation in Figure 33) condition. For participants in Group B, the order was the opposite. Participants were asked to think-aloud during the 'no Storytelling explainer' condition. In the Storytelling explainer condition, participants' attention was at selected times directed to the Storytelling HMI and participants were invited to take notes on the paper instead of thinking aloud. A Storytelling element (video) was triggered by an event occurring on the UATM HMI. When this happened, the participant was guided with a light up effect to look at the second screen with a Storytelling element (video) on the Storytelling HMI. It described something related to the event in the UATM HMI or DUC backend process that had triggered the Storytelling element. There were four Storytelling elements. Element 1 was an introductory part with terminology to UAM. Element 2 was DUC introduction - what could DUC do, what was DUC and human responsibility. Element 3 explained the sick passenger on board case and air taxi operations with sensors for passengers. Element 4 explained about the DUC's backend process (decision making process) in addressing the scenario with the sick passenger on the Air Taxi. Meanwhile, it was timed in the way that nothing important was happening on the UATM HMI.

After each condition (Storytelling explainer vs no Storytelling explainer), participants were asked to answer a questionnaire and a debriefing. The DUC questionnaire and teamwork debriefing were used for the 'no Storytelling explainer' condition. The Storytelling questionnaire and debriefing were used for the 'Storytelling explainer' condition.

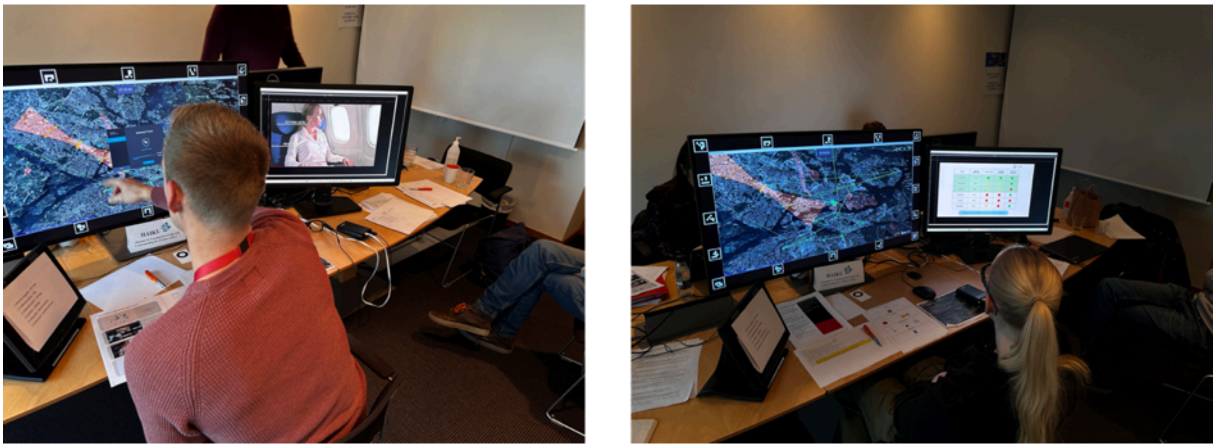


Figure 35. UAM Coordinator working position with UATM HMI on the left, and Storytelling explainer HMI on the right.

5.6 UC#3 Validation 1 Results

Analysis of results are ongoing, but some preliminary results can be provided.

- The DUC coordinating an opening of geofence in the simulation appeared as an unexpected event that raised a discussion on the roles and responsibilities of the DUC and UAM Coordinator.
- Local knowledge appeared as an important prerequisite for the role of the UAM Coordinator - Several participants were not familiar with the Stockholm area and could not orientate between locations on the map on the UAM HMI sufficiently.
- Valuable input to all three concepts
- Variation in engagement among participants provided insights to operators' needs for Human-IA Teaming aspects with respect to the roles and responsibilities of DUC and UAM Coordinator, communication, coordination, and explanations. There were several instances where the operator wanted to know more about what was happening in the simulation (what DUC was doing) and how DUC was reasoning when providing suggestions on how to solve Event 1 and Event 2.
- UAM: Separation distances could be reduced. The simulation used separation distances to approximately down to 200 metres.
- UAM: Traffic count not demanding – in future simulations we should increase the traffic count and other situations (i.e., ground events affecting the U-space) to increase the complexity.

- UAM Coordinator: Different from ATCO, more supervision
- DUC: Voice and pop-up windows worked well.

DUC Questionnaire

Participants' responses to the DUC questionnaire are provided in Table 13.

Table 13. DUC questionnaire responses. Items 1-10; 12-14 had a scale from 1-disagree highly – 7-agree highly. Item 11 had a scale from 1-Not at all – 6-exactly the same. For the acceptance scale, item 15-23 had a 1-5 scale with the terms (e.g., useful/useless) at opposite ends.

Nr	*Item*	*Median*	*Min-Max*
1	I accepted DUC's recommendation in the sick passenger rerouting situation without inspecting the situation	5	1-7
2	DUC was very helpful in handling the sick passenger rerouting situation.	7	4-7
3	I would like to know how DUC reasoned to derive the recommendation in the sick passenger rerouting situation.	7	2-7
4	I accepted DUC's recommendation in the delivery drone rerouting situation without inspecting the situation.	6	1-7
5	DUC was very helpful in handling the delivery drone rerouting situation.	7	5-7
6	I would like to know how DUC reasoned to derive the recommendation in the delivery drone rerouting situation.	5.5	2-7
7	I felt comfortable working with the DUC.	5	4-7
8	I would like to have more control over DUC.	3.5	2-6
9	A system like DUC would make my job less rewarding.	6	2-7
10	A system like DUC would allow me to focus on important aspects.	6	5-7
11	How similar is the role of the UAM Coordinator to that of an Air Traffic Controller?	2	1-4
12	DUC lowered my workload.	7	6-7
13	It felt like DUC and I worked together as a team, where DUC was my team member.	4	3-7
14	I felt that DUC was a support system or tool that I could use.	6.5	4-7
15	Acceptance scale: Please tick a box on every line. *I find the DUC…* *useful/useless*	1	1-3
16	*pleasant/unpleasant*	2.5	1-4
17	*bad/good*	4	3-5
18	*nice/annoying*	3	1-4
19	*effective/superfluous*	2	1-2
20	*irritating/likeable*	3.5	2-5
21	*assisting/worthless*	1	1-3
22	*undesirable/desirable*	4	3-5

23	*raising alertness/sleep-inducing*	3	1-5

Responses to the DUC questionnaire showed that participants consider the UAM Coordinator role different from that of an ATCO. The UAM Coordinator role is considered more of a supervisory role.

An analysis on participants' responses depending on which group they were in (Storytelling condition first or last) indicates that the Storytelling condition had small to no effects on most DUC questionnaire items. For most items, the difference between median values, between the groups, was in the order of 1 or less. However, for three items, the difference was more than 2. For item 3 (I would like to know how DUC reasoned to derive the recommendation in the sick passenger rerouting situation), the group with the Storytelling condition agree more with the statement (Median = 7) compared to the group that received the Storytelling condition last (Median = 5). In other words, the Storytelling condition first group was more interested in knowing how the DUC derived the recommendation. A similar pattern was seen for item 4 and 6. For item 4 (I accepted DUC's recommendation in the delivery drone rerouting situation without inspecting the situation) the Storytelling condition first group agreed more with the statement (Median = 6.5 compared to 4). For item 6 (I would like to know how DUC reasoned to derive the recommendation in the delivery drone rerouting situation) the Storytelling condition first group agreed more with the statement (Median = 7 compared to 3.5). These findings are noteworthy because the Storytelling condition first group had received explanations in terms of the Storytelling elements for how the DUC works and how it derives recommendations on which hospital to divert to. The explanations provided by the Storytelling elements can be considered generic in that they do not cover every pop-up message and process on the HMI screen, although the final storytelling element provides an explanation for the specific recommendation made by DUC in the simulation in emergency situations such as the medical diversion Event. Therefore, it is possible that participants in the Storytelling condition desired more specific information about every DUC generated pop-up window. In contrast, the No Storytelling condition group had not been provided with the Storytelling elements at the time of answering the DUC questionnaire.

Storytelling

Storytelling elements were enlightening, clear, and appropriate level of detail, especially for the group that had it in the second run.

- Analysis of results is ongoing, but some preliminary results can be provided.
- Looking at the storytelling explanations as a future permanent function it should rather be provided on requests from the operator, rather than "always on."
- Stories helped with a trust and understanding of the DUC decision making process. Especially the part of storytelling when the choice of the hospitals was explained in detail helped to trust the system more and make a choice on the pop up without guessing but with confidence.
- Storytelling elements 3 (air taxi sensors) and 4 (DUC choosing three hospitals) were mentioned as the most thought provoking, idea triggering and explanatory.
- The approach of storytelling was mentioned as advantageous, suitable for the role of the UAM Coordinator instead of using long manuals. Although sound narration could be optional, subtitles could be added.

- Feedback from participants was that DUC voice and pop-up windows worked better with storytelling explanations than no explanations.

Participants' responses to the Storytelling questionnaire are provided in Table 14.

Table 14. Storytelling questionnaire responses. Items had a scale from 1 to 5, where 1 is negative, 5 is positive and 3 is neutral. The table contains overall median based on all results, min-max based on all results and medians for two groups. Group A contains p

Nr	*Item*	*Median*	*Min-Max*	*Median Group A*	*Median Group B*
1	To what extent do you think that you collaborated with the DUC to solve the situations in the simulation? (From 'Not at all' to 'Very')	4	2-5	4	3.5
2	To what extent do you feel the explaining stories contributed to this interaction? (From 'Not at all' to 'Very')	4	3-5	4.5	4
3	The explaining stories made DUC (From 'More Confusing' to 'More understandable' where 3 is neutral)	4.5	4-5	4.5	4.5
4	The explaining stories made my trust in DUC (From 'Decrease' to 'Increase' where 3 is neutral)	3.5	3-5	4	3
5	DUC managing the sick passenger re-routing made me (From 'uncomfortable' to 'comfortable' where 3 is neutral)	5	4-5	5	5
6	My degree of control of the U-Space traffic made me feel (From 'uncomfortable' to 'comfortable' where 3 is neutral)	3	2-5	3.5	3
7	I feel that the story 1 explaining stories were useless/useful	5	4-5	5	5
8	I feel that the story 2 explaining stories were useless/useful	5	4-5	4.5	5
9	I feel that the story 3 explaining stories were useless/useful	5	4-5	5	5
10	I feel that the story 4 explaining stories were useless/useful	5	4-5	5	5
11	Mark on a scale of 1 to 5, how would you rate your overall experience with the explaining stories: *unfriendly/friendly*	5	4-5	4.5	5
12	*boring/exciting*	3.5	2-5	3	4
13	*confusing/clear*	4.5	4-5	4.5	4.5
14	*not interesting/interesting*	4	3-5	4	4
15	*unpredictable/predictable*	3.5	3-5	3	4
16	*complicated/easy*	5	4-5	4	5
17	*slow/fast*	3	1-5	3.5	3

18	*demotivating/motivating*	4	2-5	4.5	3.5
19	*doesn't meet expectations/ meet expectations*	4.5	3-5	4	4.5

Based on the results of the questionnaire it is possible to sum up that storytelling as an explaining approach was suitable for the specific case. For the Item 2 participants agreed on the statement that stories contributed to the interaction with DUC. For the item 3 participants also agreed on the statement that explaining storytelling made DUC more understandable. Item 4 'the explaining stories made my trust in DUC (from 'Decrease' to 'Increase' where 3 is neutral)' shows overall neutral result, Median= 3.5, although it is possible to see a differentiation in the group's answer. Group A, who did the first round with storytelling, agreed on the trust increase (Median= 4), but group B was neutral (Median= 3). The item 5 management of sick passenger re-route by DUC received after storytelling was perceived as a comfortable action (Median= 5). All 4 storytelling elements were perceived as useful (Median= 5). The overall experience with explaining stories was friendly (5), clear (4.5), interesting (4), easy (5), motivating (4) and meeting expectations (4.5). Slow/fast, boring/exciting, predictable/unpredictable were rated as neutral experiences. Statement questions were rated 0.5-1 more in group A than in group B which could suggest that having storytelling from the beginning can be to some extent advantageous for understanding, trust and collaboration between user and DUC.

Validation objectives (OBJ) and success criteria (CTR)

With respect to the validation objectives and success criteria, the preliminary results provide some answers. The **first objective** was to **assess the operational feasibility and acceptability of the concepts** (UAM, DUC, UAM Coordinator). Our results indicate that participants overall were positive to the concepts with respect to the quality of service, safety, and workload (CTR1). However, participants noted that the UAM Coordinator role was very different from that of ATCO, and that it involved more supervisory control that could be boring. The DUC was found capable of managing traffic situations in an acceptable way (CTR 2), although there was some confusion about DUCs boundaries of operations. The HMI interface was found user-friendly (CTR 3) but somewhat limiting with respect to interaction (little possibilities) and ability to access information (e.g., about aircraft).

The **second objective** was to **assess the tasks and operating methods of the UAM Coordinator**. Preliminary results indicate that the specification of the UAM Coordinators' operating methods was partly unclear (CTR 4). Some participants were more engaged and active during the simulation, e.g., by contacting external stakeholders to follow up on situations that emerged in the scenario. There is a need to more clearly establish the UAM Coordinator's operating methods.

The **third objective** was to **assess the UAM Coordinator timeliness of actions, workload, situational awareness, trust, and acceptability**. Findings indicate that participants could perform tasks in an accurate, efficient, and timely manner. However, the scenario was not perceived as complex or demanding, and there were not many tasks to be conducted given how automated the environment was (CTR 5). Overall, workload was considered low and acceptable (CTR 6). Participant's situation awareness could be improved (CTR 7). Several participants were not familiar with the Stockholm map and had difficulties to spatially orientate in the map. This shows that local knowledge is very important to the UAM Coordinator role. Participants also expressed some confusion about what the DUC was doing. The success criteria (CTR 8) addressing the acceptable level of trust and acceptability in the DUC concept was not addressed. This will be addressed in VAL 2.

The **fourth objective** was to **assess the effectiveness of the DUC in supporting the UAM Coordinator when a flight needs to deviate from its original route**. In general, participants considered the information provided by the DUC to be adequate for the situation (CTR 9). The perceived safety was not explored. DUC was also perceived capable of providing information about the most suitable route, considering traffic, ground activity/availability, distance, and airspace restrictions (CTR 10). In addition, DUC considered vehicle endurance, time to destination alternatives, and waiting time (for medical emergency situations).

The **fifth and final validation objective** was to **assess the DUC HMI interface and information requirements**. The UATM HMI received overall positive feedback from participants, although they requested more interactive possibilities (CTR 11). Tied to this, several suggestions were made on how to improve the provision of information (CTR 12), e.g., by providing flight plan data on vehicles.

5.7 UC#3 Validation 1 conclusion

Upon completion of VAL1, these are the main lessons learned that point to new design directions:

Table 15. UC#3 lessons learned

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
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 windows were appreciated by participants, but at times they	To achieve closed loop communication, DUC must be able to verify that the	Implement a function for the human to verify messages from DUC. This can be achieved by “check” buttons, a verbal

appeared to miss a message from DUC (either not hearing it or not looking at the pop-up window.	human has acknowledged/received the information provided.	repeat/check function, and supported by DUC monitoring the human's attention through the use of eye tracking.

5.8 UC#3 Appendix

Concept development

The UAM Use Case proposes a UAM Coordinator working position as envisioned in 2030. At the time of this report, there is no current UAM or UATM in operation. Therefore, the first step was to determine a concept of operations for UAM. The developed UAM concept was based on that of the U-space concept of operations proposed by the European research project: Concept of Operation for EuRopean UTM Systems - Extension for Urban Air Mobility (SESAR JU, 2022).

Urban Air Mobility (UAM) can be defined as the safe and efficient air traffic operations in a metropolitan area for manned and unmanned aircraft systems. As a starting point, there is a need for defining the UAM ConOps in which the use case and scenarios are to be realised. The HAIKU UAM use case has explored several existing UAM ConOps.

Although all represent a work in progress, none are complete, they narrow down the likely path towards a future UAM environment. The UAM use case set out to identify a suitable ConOps to lean on in the development of the digital assistant capabilities. The following ConOps were reviewed:

- CORUS XUAM (SESAR JU, 2022)
- EVE Air Mobility (EVE Air Mobility, 2022)
- FAA (FAA, 2020)
- Boeing (Boeing, 2022)
- Swiss U-space (Federal Office of Civil Aviation (FOCA), 2020)

Table 16 details the topics compared.

Table 16. UAM ConOps comparison

Topic	CORUS v4.0	EVE, Rio	FAA UTM v2.0	BOEING	Swiss U-space
USSP & CIS	Yes	Yes (UASP)	Yes (USS, PSU)	Yes (TSP)	Yes (USP)
Airspace	U-space airspaces / volumes	Defined. UAM corridors	Below 400 ft AGL in class G	No novel airspace concepts required. Class G, E , D, C and B / UAM routes	N/A
Flight Rules	UFR	N/A	To be created	Instrument procedures (RNP, IFP, IAP, SID etc.)	N/A
Flight planning	Yes (U-plan)	Yes	Yes (Operation Plan)	Automated within a UAM corridor environment	Yes
Separation services	Strategic/ tactical conflict advisory/ resolution	Tactical separation allocated to UAM operators, PIC and UASPs	Responsibility of UTM Operators	Automated traffic management within a UAM corridor environment	Yes
UAS/UAM operator	Yes	Yes (UTM Operator)	Yes (UTM Operator)	Multi-vehicle Supervisors	Yes
UAM coordinator	N/A	*N/A *in Miami ConOps, humans will still oversee the operation and be able to assist when necessary.	N/A	ATC	N/A

Clearances	Automated within U-space airspace (clearance provision to be defined)	N/A	N/A	Digital clearance delivery. (ATC issues take-off/ landing clearance (then monitor and instruct, if needed)	N/A
Communication	Verbal, textual graphic, (collaborative interface with ATCOs introduced in U3)	Voice / datalink	A distributed information network, and not between pilots and air traffic controllers via voice	Automated within a UAM corridor environment	A distributed network of digital, highly automated systems and not between pilots and controllers via voice
Phases of flight	1. Strategic – long term 2. Strategic – pre-flight 3. Pre-tactical 4. Tactical 5. Post-flight	1. Pre-flight 2. Departure 3. En route 4. Approach and landing 5. Post-flight	1. Planning 2. In-flight - Departure - En Route - Arrival 3. Post operations	1. Flight Planning 2. Pre-flight 3. Departure and Beginning of Flight 4. Inflight 5. Approach and Landing 6. Post-flight	N/A
Vertiport surface control	N/A	Yes	N/A	Vertiport managers	N/A
Contingency /Emergency Management	A service of U-space	N/A *Miami ConOps – assisted by UATM Provider	Notify to and assisted by USS	Allocated to onboard automation, with notification of and supervision by an MVS	Reporting functions

We have also considered relevant UAM related research projects:

- ICARUS (Integrated Common Altitude Reference system for U–space)[2]
- BUBBLES (separation management for UAS in the U-space) Bubbles Project home (https://bubbles-project.eu/)

[2] https://www.u-spaceicarus.eu/

- AURA (integration of ATM and U-space) SESAR JU (https://www.sesarju.eu/projects/aura)
- METROPOLIS 2 (A unified approach to airspace design and separation management for U-space) SESAR JU (https://www.sesarju.eu/projects/Metropolis2)
- USEPE (U-space Separation in Europe) USEPE (https://usepe.eu/)
- DACUS (Demand and Capacity Optimisation in U-space) DACUS (https://dacus-research.eu/)

The HAIKU UAM ConOps is based on the CORUS-XUAM ConOps. The CORUS-XUAM ConOpswas found the most comprehensive and reflects the focus, at the time, for the UAM ConOpsto be adopted in Europe. CORUS is a Horizon 2020 project funded by the SESAR Joint Undertaking that has produced four iterations of U-space ConOps, starting in 2017. The fourth iteration focuses on user needs of UAM. According to CORUS-XUAM, the term UAM is defined as air operations above urban areas (can be part of the flight) in U-space airspace where tactical separation is needed to cope with high traffic density to ensure separation. Traffic is expected to be mixed, including aircraft with limited range and incapable of flying IFR or VFR. The CORUS-XUAM operational UAM environment, with stakeholders using U-space services, is conceived in terms of three general operation types:

- Surveillance operations: mission oriented operations focusing on monitoring larger geographical areas from higher altitudes in the very low-level airspace. Such missions include UAS operations for public safety or security. E.g., Aerial mapping.
- Inspection operations: mission oriented operations requiring a close approach to a point of interest. Inspection operations are foreseen to operate within smaller confined areas. E.g., Architectural photography, power line inspection, construction site inspections.
- Transport operations: goods or passengers. E.g. Remotely piloted Air taxi operations, pharmaceutical delivery, goods delivery.

CORUS-XUAM ConOps specifies several U-space airspace volumes in controlled or uncontrolled airspace. In uncontrolled airspace, UAS operations are performed at very low levels (VLL: from GND to 500 ft AGL). The U-space airspace provides mandatory services subset into “Volumes”

- X: No conflict resolution service is offered Avoiding conflict is the responsibility of Pilot / Operator, unassisted by U-space.
- Y: Pre-flight (strategic) conflict resolution is mandatory
- Z: Pre-flight (strategic) conflict resolution and in-flight (tactical) conflict resolution are mandatory.

The Z volume is further divided into:

- Za in controlled control zones (CTR) around e.g., airports for which ATCO manages all the traffic. ATC is expected to benefit from U-space services in managing UAS.
- Zu overhead urban areas in which U-space will provide a tactical conflict resolution service.
- Zz overhead urban areas in which U-space will provide a tactical conflict advisory service.

For higher operations (e.g., inter-cities passenger/cargo flights), corridors will be provided and published in AIP. In more densely occupied U-space airspace, tactical conflict resolution is remotely offered. UAS traffic in ATC controlled areas will be controlled by ATC through U-space. A dynamic capacity management service will be needed to match the capacity and traffic demand.

Some U-space airspaces with tactical services will accommodate remotely piloted flights according to U-space Flight Rules (UFR).

The following procedure is proposed in CORUS-XUAM for U-space services for scheduled passenger flights: In Pre-flight, the UAS operator submits a U-plan to the Flight Authorization service. The UAS operator is informed that the flight is conflict free at reasonable time to act (RTTA). If conflict, strategic conflicts are generally solved by speed and route changes. When in-flight, the U-Space Service Provider (USSP) activates the U-plan and commences the tactical services and monitoring required for the airspace (in Zu, the pilot receives tactical separation instructions).

The CORUS-XUAM ConOps does not envision or explore the needs for UAM service providers, UAM management, or digital assistants for UAM. However, the ConOps does specify essential U-space services to be provided to U-space users:

- Network Identification,
- Geo-awareness,
- Flight authorization,
- Traffic information,
- Weather information, and
- Conformance monitoring.

The role of the USSP is central to the U-space ConOps and could comprise the following as part of UTM (first four mandatory, the two last optional):

- Manage (activate/deactivate) U-space airspace in controlled airspace.
- Authorise vertiport landings and take-offs according to vertiport slots and infrastructure.
- Flight planning and flight authorization services to UAS operators.
- Conformance monitoring
- Support detect and avoid (DAA) in e.g., Zu airspace (resolution service) or Zz airspace (resolution advisory only) , note that it is assumed that aircraft will have advanced sense and avoid technologies that make conventional ATC conflict detection and resolution CD&R redundant. Nonetheless, a USSP might be required to provide CD&R services in some situations.
- Coordinate traffic management with ATC through a collaborative interface. The CORUS-XUAM ConOps also discuss the “collaborative interface with ATC” as “a service offering communication between ATC and the appropriate representative of a drone flight… or in some cases the USSP.”
- Vertical Alert and Information Service. The USSP could provide navigation information to UAS and manned flights (e.g., general aviation) on ground collision risks.
- Emergency management

A key concern with UAM is how to cope with emergency situations and the limited possibilities for emergency landings. The urban landscape poses several risks to both ground and air. Emergency management is a critical service to be provided by the USSP. It incorporates:

- Assistance to UAS operator/pilot experiencing emergency (e.g., report, action proposal, contingency plans, emergency response plan)
- Inform and communicate with other stakeholders of emergency, such as other traffic and emergency response (on ground and in air).
- Emergency procedures (e.g., configure dynamic safety boundaries, change flight prioritisation, plan and coordinate emergency routings).

Flights in U-space can be classified in two priority categories according to EU regulation 2021/664 article 10, paragraph 8: When processing UAS flight authorization requests, the U-space service providers shall give priority to UAS conducting special operations as referred to in Article 4 of Implementing Regulation (EU) No 923/2012.

In turn, Article 4 of EU regulation 923/2012 (SERA, the Standard European Rules of the Air) lists special operations that may be granted exemptions:

1) police and customs missions;
2) traffic surveillance and pursuit missions;
3) environmental control missions conducted by, or on behalf of public authorities;
4) search and rescue;
5) medical flights; (HAIKU use case)
6) evacuations;
7) fire fighting;
8) exemptions required to ensure the security of flights by heads of State, Ministers and comparable State functionaries.

For HAIKU, the special operation of medical flights is used as an example for further exploration.

UAM ConOps and research on UAM or similar, generally assume that the UAM system will be fully automated. Detection and avoidance (DAA) are envisioned to be automated: vehicles will have their own sense and avoid solutions. There has been very limited focus on exploring the human role in envisioned UAM landscapes. UTM is envisioned to be autonomous involving as least human as possible. If anything, the human is envisioned to oversee the operation and be able to assist, when necessary (supervising role), or the human could be a fallback system when automation fails. HAIKU sees a need for humans engaging with the UAM system to accommodate the link between ground and airborne activities.

The HAIKU use case workshop identified the following key control tasks as part of the UAM management ConOps:

1) Coordinate search and rescue activities (there is an operations priority list in 664 that covers this – JRCC)
    a) Analyse city areas for this activity.
    b) Prioritise the size of the area when others want to fly/operate there.
    c) Manage other traffic so that they are out of the emergency area.
    d) Does every emergency trigger a “no flight” zone at all altitudes?
2) Prioritise between space and stability. Need to set flight priority?
    a) Need for new priority rules and thresholds (limit values)
    b) Prioritise different services and operations.
    c) Delay someone more to reduce overall/total delay of all/in the system.

   d) Blue light unit and military units
3) Issue clearances.
   a) Same language – analogue or digital.
   b) Enter controller or uncontrolled airspace.
   c) Priority of clearances.
   d) Democratic principles – public transparency (the principle of publicity) and fair treatment (fairness/justice)
4) Monitor deviations.
   a) Judge deviation – how severe or difficult. Activate safety net tools.
   b) Requirement for automatic safety measures or manual.
   c) Calibration of different limit values/thresholds for deviations, e.g., separation criteria (distance).
5) The effect of the city on UAM
   a) Fires that create restricted areas (no-fly zones).
   b) Confirm and activate these areas as restricted.
   c) Activate areas (open up) when no-go activity is finished.
   d) Check non-conformance alert (vehicles violating no-go zone).

In an interview with UAM expert Liza Chi Hang Josias, ecosystem & stakeholder engagement specialist from Eve Air Mobility, the following key important tasks were identified in requirements for human supervision of UAM:

- Flow management: monitoring system, making sure flow is working as needed.
- Responding to changes in the system, e.g. vertiport going offline.
- Managing non-normal situations.
- Requests from other stakeholders, such as ANSP or cities.
- Coordinating between ANSP, cities, done operators, first response services and emergency services.
- Role of the human in managing UAM is closer to road traffic managers than ATOCs.

In HAIKU we challenge the fully-automated view as we envision a need for human involvement in safely and efficiently managing the airspace above cities and coordinating UAM with ground activities. We envision this human role as a UAM Coordinator. The role and responsibility of the UAM Coordinator would be to provide U-Space services to UAS and UAM operators, including geo-awareness services (e.g. configuration of airspace constraints and temporary restrictions), flight authorization (managing and authorising flight plans), and traffic monitoring. The key purpose of the UAM Coordinator is, as the name suggests, to coordinate between stakeholders involved in the UAM ecosystem. This is in contrast to ATCOs who manage traffic. The UAM Coordinator would probably be an actor employed by the local USSP. In HAIKU we envision that each city would have its own local USSP, where a single or multiple UAM coordinators could be responsible for the airspace overhead the city (e.g., if there are multiple USSPs within a city). The UAM coordinator has some resemblance to ATCOs in that both provide a service to aircraft. A difference, however, is that aircraft in the U-space, in contrast to controlled airspace in ATM, are envisioned to operate autonomously to solve route monitoring and CD&R without the need for human intervention. As such, a more suitable comparison might be with road traffic management operators. In road traffic management, cars operate ‘autonomously’. Instead, the road traffic management operators focus on flow management, managing constraints (e.g., closing roads), and reacting to emergency situations. Similarly, the role of the UAM Coordinator is envisioned as a requirement for managing the flow of aircraft, geo-fence provisions, reacting to emergency situations, and coordinating dependencies between ground-based activities and aircraft operations.

Note that many of the USSP services are assumed to be highly automated, but some are likely to require human involvement to a higher degree due to the uncertain and dynamic nature of the service. One such critical service focused on in HAIKU is emergency services.

We envision the need for a UAM Coordinator role to be especially important in the growth and development of UAM and U-space services (i.e., as U-space operations transitions U2 - initial services, to U3 - advanced service, toward U4 - full services). As such, the target for c0f2f3 activities is 2030. However, as UAM approaches full services (U4), the need for a UAM Coordinator role may diminish with increased autonomy.

6. Use Case #4 – Digital and Remote Tower

6.1 UC#4 Validation 1 Plan Summary

The VAL1 plan for Use Case 4 is designed to validate the Intelligent Sequencer Assistant (ISA) concept at the current stage of development. 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.

VAL1 aims to investigate the following Key Performance Areas (KPAs):

Table 17. UC#4 performance targets

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).

Starting from these KPAs, VAL1 assesses the following R&D objectives:

Table 18. UC#4 R&D objectives

OBJ-ID	Validation Objective	Success Criteria ID	Success Criteria
UC4-OBJ-01	To assess the Operational Feasibility Acceptability of the solution from the ATC perspective in nominal conditions.	UC4-CTR-01	The solution is considered Operationally Feasible and 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	The ATCOs give positive feedback on the current status of the HMI and the explainability provided.

6.2 Deviation from Initial Validation Plan

The original Validation Plan has undergone slight modifications to accommodate new developments in the AI's design process:

- UC4-OBJ-04 has been added to the R&D objectives to validate the current state of the Human-Machine Interface (HMI) due to the detailed interface design produced in the previous months.
- The concept of conformance has been introduced to the Key Performance Area (KPA) of Human Performance. This addition aims to facilitate a comparison between ISA's solutions and those of Air Traffic Control Officers (ATCOs) to determine if the solutions align with ATCO practices and mental models.
- The original plan to structure exercise scenarios in the simulator with varying difficulty levels has been abandoned as it was deemed non-contributory to the experiment assessing validity in nominal conditions. Instead, two scenarios with constant difficulty levels were designed to be more realistic and aligned with actual situations where ATCO might use ISA.

6.3 UC#4 Demonstrator and Assistant Features

For VAL1, the ISA prototype can provide sequence solutions to address specific ATC scenarios. At this stage of development, the sequence provided cannot be generated in real-time and there is not a high-fidelity front-end available that users can interact with. However, even without interaction and with sequences generated asynchronously, they can still be compared to those of real ATCOs because the volume of traffic and the aircraft provided in the exercises are the same. Thus, valuable results can still be obtained from this validation that can be used to inform the next stages of the project. The current features investigated during VAL1 are the following: the ISA Engine and the HMI.

ISA Engine

To provide sequence solutions, ISA uses an engine that relies on data obtained from simulation exercises within ATCO training, incorporating both simulator-generated data and trainee activity records. The ongoing work in this UC focused on designing the AI optimisation engine components and underlying algorithms necessary to deliver the expected sequence. The data sink infrastructure is developed and set up, with testing and integration activities underway. In terms of ML/DL work, the following have been implemented:

- A service that computes the estimated arrival time for arrival flights on the active runway (direction) and the initial points of the incoming flight's trajectory. Two types of models have been trained with very similar performance, specifically a linear regression model and a feed-forward neural network with one hidden layer. The aircraft's current position, true airspeed and track are used as input for the aforementioned models, whose output is the remaining time in seconds until the aircraft lands. As the ML model developed requires more training data to be accurate enough, for the VAL1 phase we used ETA information extracted directly from the simulation data in order to ensure that the input to the sequence calculation, which is the core part of the system, was kept as close as possible to the information available to the ATCO during the simulation. In this way we could obtain comparable results for the two sequences, the one by ISA and the one by ATCO.
- A service that configures and applies an optimisation algorithm using the current set of active aircraft, i.e. the aircraft which the controller needs to consider for sequencing. The problem is formulated as a MILP problem. Its input includes the estimated arrival time for arrival flights, the EOBT/CTOT for departure flights together with the tolerance windows, the wake turbulence category and the flight rules category (instrument or visual) for each aircraft, and the estimated taxi times. The model returns the flight sequence (including arrivals and departures) that minimises the total runway usage time, i.e. the sequence that ensures the last flight of the sequence will depart/land as soon as possible.

The aforementioned models are configured and trained using 172 simulation exercises, each one containing multiple arrivals and departures. It should be noted that for the first phase, edge cases (e.g. trajectories with irregular movements and training flights) were excluded from the analysis.

Human-Machine Interface

On the front-end side, the HMI has been designed as a mock-up and it was not integrated with the backend at the current stage of the project (Figure 36). However, the HMI mock-up is mature enough to be assessed during VAL1 to gather initial feedback before its full development.

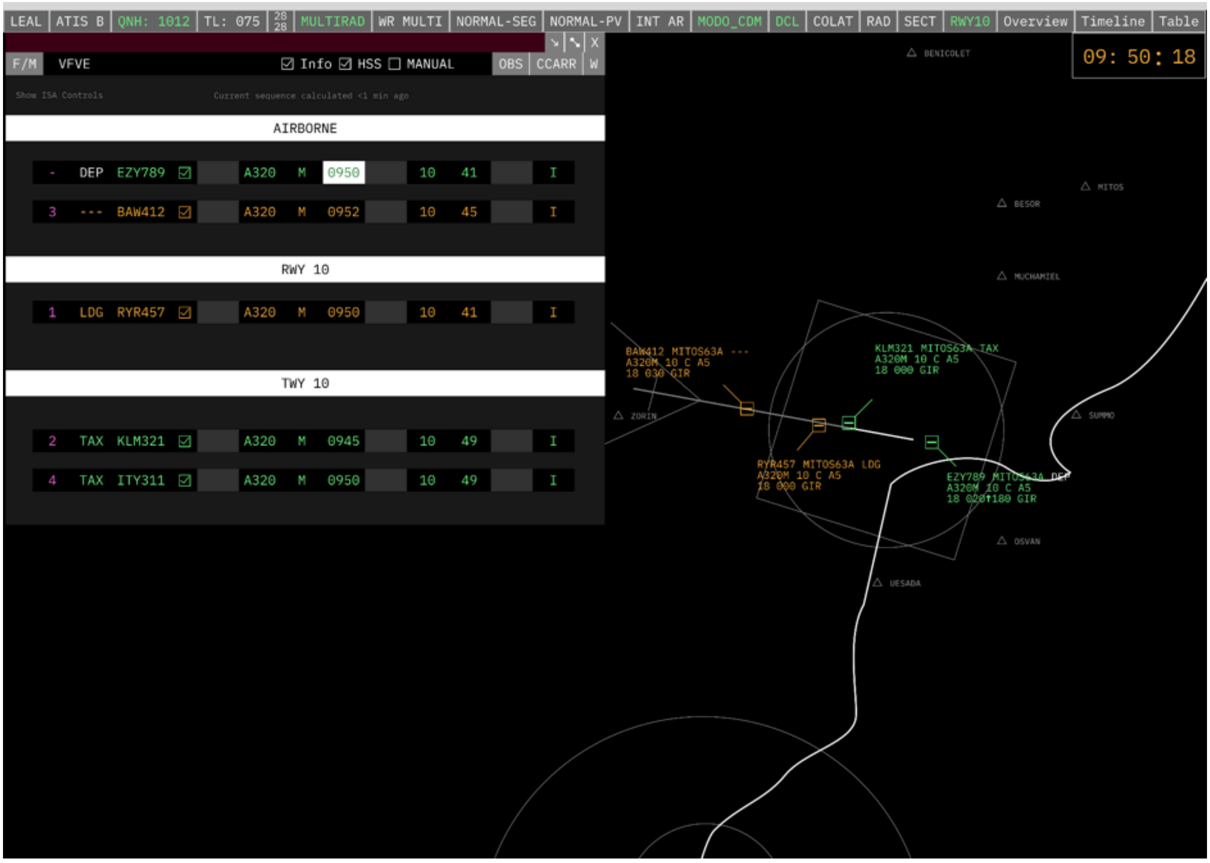


Figure 36. Human-Machine interface of the ISA prototype

The current HMI is designed to be integrated in a non-disruptive manner into Alicante's current system. It provides the sequence in real-time to controllers, showcasing relevant changes and providing varying levels of information based on the CLT framework outlined in D5.1. The CLT levels (1 through 6) are implemented as follows:

- (CLT 1) *Overview of Main KPIs*: This panel displays key performance indicators, such as departures, arrivals, safety events, and current average runway usage buffer in the next 60 minutes (Figure 37). Users can select different time frames from a dropdown menu to observe expected KPI evolution.

Figure 37. ISA Key performance indicator panel

- (CLT 2) *Ordered Sequence Display*: The electronic strip of each aircraft displays a number that corresponds to their place in the sequence (Figure 38). Hovering the mouse on the numbers reveals the explanation for each specific position.

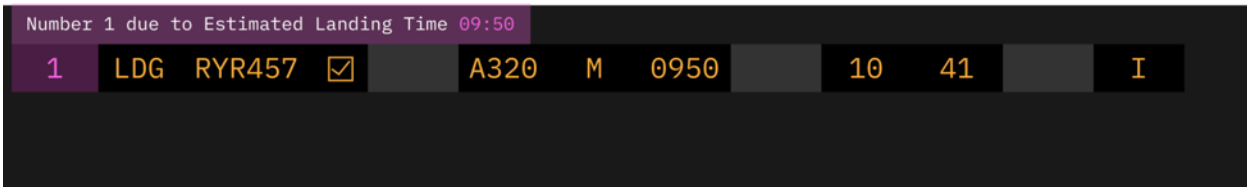


Figure 38. ISA ordered sequence display

- (CLT 3) *Sequence Change*: When ISA detects an event, the sequence may require adjustment. ISA highlights the electronic strips of involved aircraft, using a blinking effect for easy identification by ATCOs. After completion, hovering over the newly changed number provides a concise one-line explanation, with the option for a more detailed explanation through a right-click "on-demand" panel (currently a work in progress) (Figure 39).

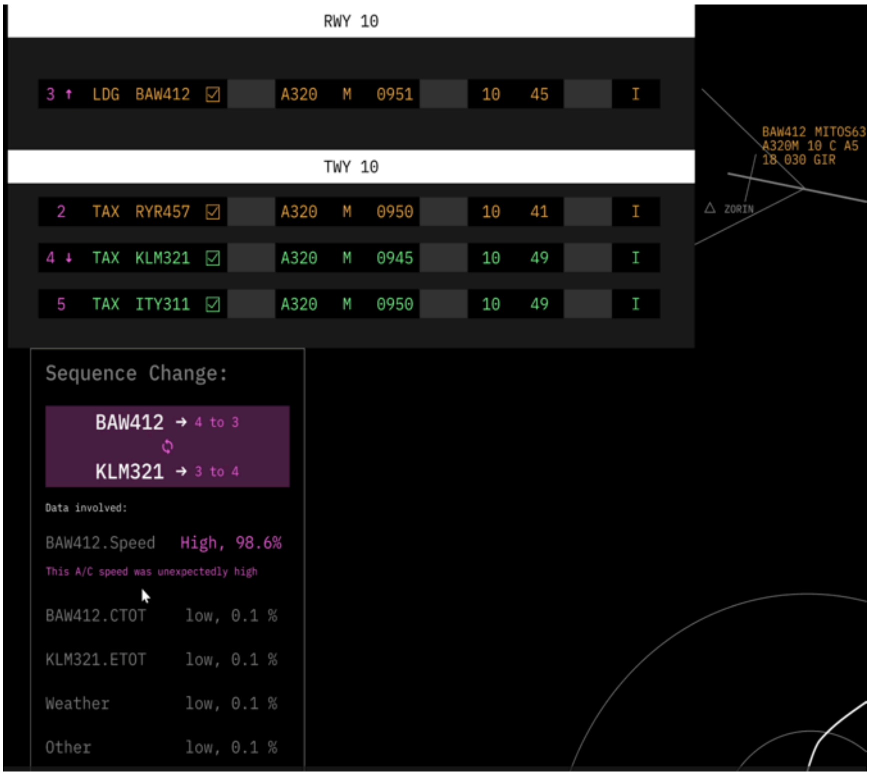


Figure 39. ISA on-demand panel

- (CLT 4) *Timeline*: Intended for post-operation use, the timeline presents a timeline where ATCOs can choose the level of detail (minutes, hours, days). All ISA events are displayed, allowing ATCOs to retrace steps, revisit specific events, and access detailed explanations (Figure 40).

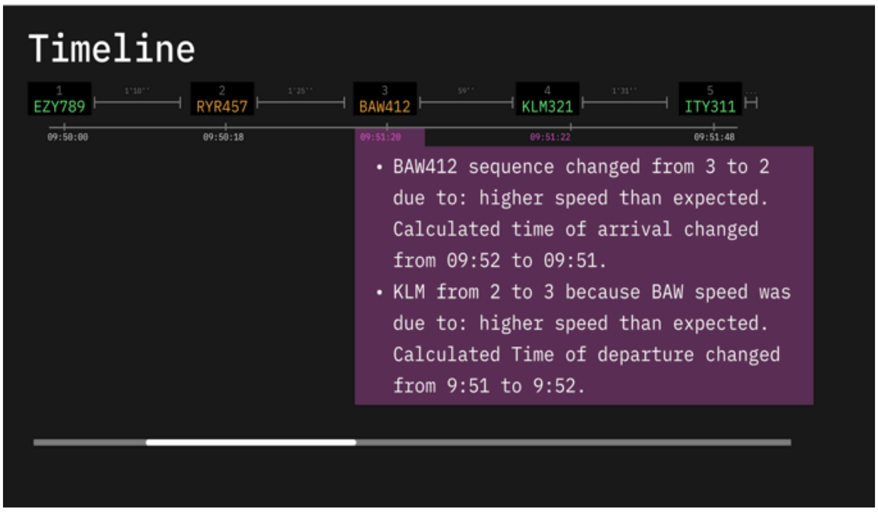


Figure 40. ISA post-operation timeline

- (CLT 5-6) *Tables with Full Descriptions*: Intended for post-operation use, data tables exhibit comprehensive information, providing high-detail explanations for all sequence changes. These tables should be used to understand the underlying reasons for operational adjustments (Figure 41).

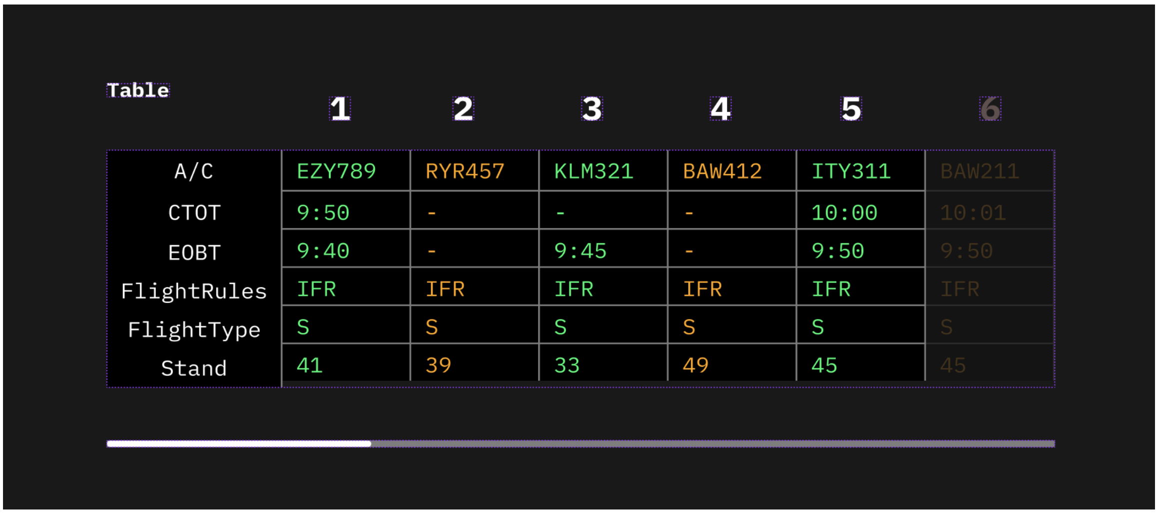


Table	1	2	3	4	5	6
A/C	EZY789	RYR457	KLM321	BAW412	ITY311	BAW211
CTOT	9:50	-	-	-	10:00	10:01
EOBT	9:40	-	9:45	-	9:50	9:50
FlightRules	IFR	IFR	IFR	IFR	IFR	IFR
FlightType	S	S	S	S	S	S
Stand	41	39	33	49	45	45

Figure 41. ISA post-operation information table

6.4 UC#4 TRL

In UC4, there are five system components related to the intelligent assistant that were explored. They are reported below, including their initial TRL:

- First iteration of the Intelligent Sequence Assistant (ISA) in a simulator. That included the following subsystems:
  - Core data processing module - TRL 4
  - ML module for ETA calculation - TRL 3
  - Sequence calculation module - TRL 4
  - Explanation generation module - TRL 3
- HMI (video showing a non-interactive interface): TRL 3.

In VAL1, our first ISA iteration was tested in the simulator by ATCOs. Core data processing module, ETA calculation module and sequence calculation module were tested using pre-recorded exercises, and ISA was able to provide a sequence that changed during the exercise. This is considered TRL4.

Note that the real-time suggestions were not implemented in VAL1 so the Explanation generation module could not be tested at the same level. Also, the ML module was not fully implemented so it was mostly used as an experimental proof of concept.

The IA interface was already designed based on the current real-live working position and was shown as a video following a sequence of events. Therefore, the HMI should be considered TRL 3 Experimental proof of concept as it replicates exactly how the HMI will work in the simulator and it was very valuable to demonstrate the critical functions and characteristics of the system.

6.5 UC#4 Validation 1 Activities

Scenarios

Three scenarios were designed for the validation activities, detailed as follows:

Scenario 1 - Alicante Control Tower Free Simulation

In this exercise, a realistic scenario of a working shift of an ATCO in ALC is replicated. ATCO's goal is to sequence all departures and arrivals in the most efficient way possible. A solution for this scenario by ISA is already available (done asynchronously), but the ATCO solving the exercise does not know what the solution suggested by the ISA is. The scope of the exercise is to assess the following: comparison between the ideal sequence and ATCO sequence and level of conformity, number of safety events, workload, and situational awareness (UC4-OBJ-01/02/03).

Scenario 2 - Alicante Control Tower Simulation with ISA instructions

The exercise is to be performed in a simulator that replicates the same conditions as in the ALC monoposition. Scenario 2 requires the ATCO to solve the scenario by implementing a sequence suggested by ISA. The scope of the exercise is to assess the following: Trust, quality of AI resolutions and Human-AI teaming aspect (UC4-OBJ-01/02/03).

Scenario 3 - HMI Evaluation

This part of the experiment consists of a video showing the main features of the HMI, simulating its use in a scenario where a resequencing process is triggered. The scope of the exercise is to assess how ATCOs perceive and evaluate the HMI in terms of usability and user experience, and how they evaluate the solutions from the explainability point of view (UC4-OBJ-04).

Procedure

The validation activities involved 8 (7 men, 1 female. average age = 39.1) ATCOs with different levels of experience who participated in all three scenarios. All sessions were carried out in the simulator room of Skyway in Madrid, with the software set up to replicate Alicante’s Airport layout and context. Upon consent form completion, each participant was first introduced to the HAIKU project and to UC4 research objectives.

After the initial briefing, the participants were asked to take the mono position at the simulator. When everything was set up, with pseudo pilots and technical scenarios ready, the experiment would start with Scenario 1. Each scenario would last approximately 30 minutes, with a 15-minute debrief at the end.

During the experiments, data collection involved quantitative and qualitative data. The simulator log recorded all the essential KPIs pertaining to the exercises, such as number of arrival/departures, safety events and aircraft data, though it must be highlighted that, given the asynchronous nature of the exercise (with sequences being different for each participant), those KPIs were not relevant in VAL1. On the other hand, the actual sequences produced by the participants were logged and compared with ISA's to reflect on all the possible situations in which the IA might be used.

From a qualitative perspective, there was an extensive collection of data. The following methodologies were used:

- Direct observations: the experimenters systematically observed participants' behaviours, actions, and responses during the validation activities.
- Think-aloud protocols: participants were asked to think aloud while carrying out their tasks, specifically in part 2 of the exercise.
- Semi-structured interviews: Post-activity interviews conducted conversationally, with a predetermined set of open-ended questions to investigate the main KPAs. These interviews occurred during the debriefing sessions following each activity, enriching the qualitative data with participant insights.

6.6 UC#4 Validation 1 Results

General Feedback

Overall, the participants evaluated the concept of ISA positively. P1 said "This can surely be helpful", while P8 said "This is good. This is the future that we will have to deal with". However, the participants also pointed out that ISA can be a double-edged sword in the sense that, while it can aid resolution of sequencing problems, it might also distract the controller in tight situations.

The following paragraphs detail the results concerning each R&D Objective.

Operational Feasibility and Acceptability

In terms of Operational Feasibility and Acceptability, ATCO considered the sequence provided viable and acceptable, even if provided asynchronously. P8 mentioned that the sequences provided were "logical".

Regarding Scenario 1, these are the main results:

- There were minimal differences between sequences generated by ATCO and ISA, and these would mostly be attributed to the simulator and pseudo pilots causing unforeseen variations.
- In nominal conditions, one result of the asynchronicity was that ISA seemed to take risks and act aggressively even when unnecessary. Conversely, ATCOs tend to act more conservatively. For example, in Scenario 1, there was a tight situation where an aircraft was set to arrive just as another aircraft was about to take off. All ATCOs chose to act conservatively in that situation by stopping the take-off and allowing the landing first,

whereas ISA always pushed the departure because, mathematically, there was just enough time.

- In scenarios with only two aircraft involved, ATCOs would not accept ISA's solutions because there would be no point in taking any risk.
- At the end of the exercises, ATCOs were asked to self-evaluate their stress levels during the exercise. For most of the time, all ATCOs said they were "Relaxed" or "Focused", though on those extremely tight sequences generated by the simulator some of them felt "Under pressure". These moments were identified as moments in which ISA would be more useful.

Below are the sequences produced by ATCOs in Scenario 1, compared with ISA's.

VAL1-01				
ATCO:	AMG	DATE:		16/1/2024
ISA'S SEQUENCE:		CURRENT SEQUENCE:		
1	AZA421:	1	AZA421:	22:05
2	EZY22LR:	2	EZY22LR:	22:08
3	IBE247:	3	IBE247:	22:10
4	VLG214:	4	VLG214:	22:14
5	JAF115:	5	JAF115:	22:16
6	IBE2836:	6	IBE2836:	22:18
7	TOM424:	7	TOM424:	22:22
8	ABY2511:	8	ABY2511:	22:24
9	SWT1211:	10	SWT1211:	22:30
10	AFR56C:	9	AFR56C:	22:26
11	TRA484F:	11	TRA484F:	22:31
12	RYR8ZF:	12	RYR8ZF:	22:32

VAL1-01				
ATCO:	PLG	DATE:		22/1/2024
ISA'S SEQUENCE:		CURRENT SEQUENCE:		
1	AZA421:	1	AZA421:	22:06
2	EZY22LR:	3	EZY22LR:	22:10
3	IBE247:	2	IBE247:	22:09
4	VLG214:	4	VLG214:	22:11
5	JAF115:	5	JAF115:	22:14
6	IBE2836:	6	IBE2836:	22:15
7	TOM424:	7	TOM424:	22:16
8	ABY2511:	8	ABY2511:	22:17
9	SWT1211:	10	SWT1211:	22:21
10	AFR56C:	9	AFR56C:	22:19
11	TRA484F:	11	TRA484F:	22:23
12	RYR8ZF:	12	RYR8ZF:	22:25

VAL1-01				
ATCO:	JRP	DATE:		16/1/2024
ISA'S SEQUENCE:		CURRENT SEQUENCE:		
1	AZA421:	1	AZA421:	22:07
2	EZY22LR:	2	EZY22LR:	22:09
3	IBE247:	3	IBE247:	22:11
4	VLG214:	4	VLG214:	22:13
5	JAF115:	7	JAF115:	22:23
6	IBE2836:	5	IBE2836:	22:17
7	TOM424:	8	TOM424:	22:24
8	ABY2511:	6	ABY2511:	22:20
9	SWT1211:	9	SWT1211:	22:25
10	AFR56C:	10	AFR56C:	22:28
11	TRA484F:	11	TRA484F:	22:30
12	RYR8ZF:	12	RYR8ZF:	22:31

VAL1-01				
ATCO:	IMY	DATE:		22/1/2024
ISA'S SEQUENCE:		CURRENT SEQUENCE:		
1	AZA421:	1	AZA421:	22:06
2	EZY22LR:	3	EZY22LR:	22:09
3	IBE247:	2	IBE247:	22:08
4	VLG214:	4	VLG214:	22:10
5	JAF115:	5	JAF115:	22:12
6	IBE2836:	6	IBE2836:	22:14
7	TOM424:	7	TOM424:	22:16
8	ABY2511:	8	ABY2511:	22:18
9	SWT1211:	9	SWT1211:	22:21
10	AFR56C:	12	AFR56C:	22:25
11	TRA484F:	10	TRA484F:	22:23
12	RYR8ZF:	11	RYR8ZF:	22:26

VAL1-01				
ATCO:	ATS	DATE:		17/1/2024
ISA'S SEQUENCE:		CURRENT SEQUENCE:		
1	AZA421:	1	AZA421:	22:06
2	EZY22LR:	3	EZY22LR:	22:10
3	IBE247:	2	IBE247:	22:09
4	VLG214:	4	VLG214:	22:13
5	JAF115:	6	JAF115:	22:19
6	IBE2836:	5	IBE2836:	22:18
7	TOM424:	8	TOM424:	22:25
8	ABY2511:	7	ABY2511:	22:23
9	SWT1211:	9	SWT1211:	22:26
10	AFR56C:	10	AFR56C:	22:29
11	TRA484F:	12	TRA484F:	22:32
12	RYR8ZF:	11	RYR8ZF:	22:31

VAL1-01				
ATCO:	SWM	DATE:		24/1/2024
ISA'S SEQUENCE:		CURRENT SEQUENCE:		
1	AZA421:	1	AZA421:	22:06
2	EZY22LR:	3	EZY22LR:	22:11
3	IBE247:	2	IBE247:	22:09
4	VLG214:	4	VLG214:	22:12
5	JAF115:	5	JAF115:	22:14
6	IBE2836:	6	IBE2836:	22:18
7	TOM424:	7	TOM424:	22:19
8	ABY2511:	8	ABY2511:	22:20
9	SWT1211:	9	SWT1211:	22:23
10	AFR56C:	10	AFR56C:	22:26
11	TRA484F:	12	TRA484F:	22:29
12	RYR8ZF:	11	RYR8ZF:	22:28

VAL1-01				
ATCO:	RLL	DATE:		17/1/2024
ISA'S SEQUENCE:		CURRENT SEQUENCE:		
1	AZA421:	1	AZA421:	22:06
2	EZY22LR:	3	EZY22LR:	22:10
3	IBE247:	2	IBE247:	22:09
4	VLG214:	4	VLG214:	22:14
5	JAF115:	5	JAF115:	22:18
6	IBE2836:	6	IBE2836:	22:19
7	TOM424:	8	TOM424:	22:23
8	ABY2511:	7	ABY2511:	22:22
9	SWT1211:	9	SWT1211:	22:28
10	AFR56C:	10	AFR56C:	22:29
11	TRA484F:	11	TRA484F:	22:31
12	RYR8ZF:	12	RYR8ZF:	22:32

VAL1-01				
ATCO:	JPM	DATE:		24/1/2024
ISA'S SEQUENCE:		CURRENT SEQUENCE:		
1	AZA421:	1	AZA421:	22:06
2	EZY22LR:	3	EZY22LR:	22:09
3	IBE247:	2	IBE247:	22:08
4	VLG214:	4	VLG214:	22:14
5	JAF115:	5	JAF115:	22:15
6	IBE2836:	7	IBE2836:	22:18
7	TOM424:	6	TOM424:	22:17
8	ABY2511:	8	ABY2511:	22:22
9	SWT1211:	9	SWT1211:	22:24
10	AFR56C:	10	AFR56C:	22:26
11	TRA484F:	11	TRA484F:	22:29
12	RYR8ZF:	12	RYR8ZF:	22:27

Figure 42. Sequences produced by ISA and ATCOs in scenario 1

Regarding Scenario 2:

- ATCOs implemented exactly what ISA suggested and, barring some moments caused by the pseudo pilots which led to some differences, they mostly agreed with the sequence provided.
- ATCOs understood the rationale why ISA suggested sequence changes.
- The key point for this exercise is that ATCOs mentioned that their mental picture of the situation might change from one minute to the other, and it is never fixed. So long as nothing happens, they would accept ISA's solutions, but if something unexpected happens, they must be ready to adapt immediately and change the sequence.
- Based on this, it is clear that ISA must be operating in real-time and be able to change the sequence of aircraft on the fly so that ATCOs can look at it. Without these real-time capabilities, the concept is not going to be operationally feasible.
- Just like Scenario 1, in terms of Human Performance, all ATCOs said they were "Relaxed" or "Focused", though some said they were "Under pressure" in some specific moments. Again, the situations in which they found themselves to be more pressured were due to tricky aircraft movements that required precision and very quick thinking. Given that in this exercise they already had the sequence provided by ISA, they said that having a tool to help them in these situations would be useful.

Below are the sequences produced by ATCOs in Scenario 2, compared with ISA's.

VAL1-02			
ATCO:	AMG	DATE:	16/1/2024
ISA'S SEQUENCE:		YOUR IDEAL SEQUENCE:	
1	H2315:	1	H2315
2	M642:	3	M642
3	A124:	2	A124
4	A515:	4	A515
5	R241:	5	R241
6	L989:	7	L989
7	R2JD:	6	R2JD
8	L311Y:	8	L311Y
9	X45A:	9	X45A
10	0MY:	10	05MY
11	32VM:	11	32VM

VAL1-02			
ATCO:	PLG	DATE:	22/1/2024
ISA'S SEQUENCE:		YOUR IDEAL SEQUENCE:	
1	H2315:	1	H2315
2	M642:	2	M642
3	A124:	3	A124
4	A515:	4	A515
5	R241:	5	R241
6	L989:	6	L989
7	R2JD:	7	R2JD
8	L311Y:	8	L311Y
9	X45A:	9	X45A
10	0MY:	10	05MY
11	32VM:	11	32VM

VAL1-02			
ATCO:	JRP	DATE:	16/1/2024
ISA'S SEQUENCE:		YOUR IDEAL SEQUENCE:	
1	H2315:	1	H2315
2	M642:	2	M642
3	A124:	4	A124
4	A515:	3	A515
5	R241:	5	R241
6	L989:	6	L989
7	R2JD:	7	R2JD
8	L311Y:	8	L311Y
9	X45A:	9	X45A
10	0MY:	11	05MY
11	32VM:	10	32VM

VAL1-02			
ATCO:	IMY	DATE:	22/1/2024
ISA'S SEQUENCE:		YOUR IDEAL SEQUENCE:	
1	H2315:	1	H2315
2	M642:	2	M642
3	A124:	3	A124
4	A515:	4	A515
5	R241:	5	R241
6	L989:	6	L989
7	R2JD:	7	R2JD
8	L311Y:	8	L311Y
9	X45A:	9	X45A
10	0MY:	10	05MY
11	32VM:	11	32VM

VAL1-02			
ATCO:	ATS	DATE:	17/1/2024
ISA'S SEQUENCE:		YOUR IDEAL SEQUENCE:	
1	H2315:	1	H2315
2	M642:	2	M642
3	A124:	3	A124
4	A515:	4	A515
5	R241:	5	R241
6	L989:	7	L989
7	R2JD:	6	R2JD
8	L311Y:	8	L311Y
9	X45A:	9	X45A
10	0MY:	10	05MY
11	32VM:	11	32VM

VAL1-02			
ATCO:	SWM	DATE:	24/1/2024
ISA'S SEQUENCE:		YOUR IDEAL SEQUENCE:	
1	H2315:	1	H2315
2	M642:	2	M642
3	A124:	3	A124
4	A515:	4	A515
5	R241:	5	R241
6	L989:	6	L989
7	R2JD:	7	R2JD
8	L311Y:	8	L311Y
9	X45A:	9	X45A
10	0MY:	10	05MY
11	32VM:	11	32VM

VAL1-02			
ATCO:	RLL	DATE:	17/1/2024
ISA'S SEQUENCE:		YOUR IDEAL SEQUENCE:	
1	H2315:	1	H2315
2	M642:	2	M642
3	A124:	3	A124
4	A515:	4	A515
5	R241:	5	R241
6	L989:	6	L989
7	R2JD:	7	R2JD
8	L311Y:	8	L311Y
9	X45A:	9	X45A
10	0MY:	10	05MY
11	32VM:	11	32VM

VAL1-02			
ATCO:	JPM	DATE:	24/1/2024
ISA'S SEQUENCE:		YOUR IDEAL SEQUENCE:	
1	H2315:	1	H2315
2	M642:	3	M642
3	A124:	2	A124
4	A515:	4	A515
5	R241:	6	R241
6	L989:	5	L989
7	R2JD:	7	R2JD
8	L311Y:	8	L311Y
9	X45A:	9	X45A
10	0MY:	11	05MY
11	32VM:	10	32VM

Figure 43. Sequences produced by ISA and ATCOs in scenario 2

On a general note, in the debriefing of the scenarios the following points emerged regarding ISA's feasibility and acceptability:

- If there are no tricky conflicts, ATCOs would simply not use ISA as they can do all the calculations themselves, as shown concretely in practice and highlighted by their comments in the debriefing.
- It seems that for ISA to be operationally viable, its scope must be reduced, and it should help the controllers in those tight moments where controllers need to do many calculations in their mind and are unsure of what the next steps are, rather than be active at all times.

Safety

In terms of safety, ISA is perceived:

- As a mixed bag, probably due to the ATCOs having different experiences and perceptions during the experiment. P1 said “If it gives the right sequence, it could improve the fluency of operations. But, on the other hand, an inexperienced ATCO might feel pressured to follow the sequence recklessly”. P3 said that at the moment the “concept seems to be more tuned to operational efficiency rather than safety”. P7 seemed particularly enthusiastic, outlining how “it can help you by backing up your decisions with facts in tense moments and avoid problems”.
- As “conservative”, “moderate” and “aggressive” in terms of ATC style. The participants’ comments were all over the place, and it seems as if there is no point of balance at the moment, probably due to the asynchronous nature of the exercise.

Trust

For a system such as ISA, trust is extremely important and goes hand in hand with operational performance. P2 said “If you don’t trust it, then it doesn’t make any sense”. The main takeaways from the experiments are:

- Given ISA’s good performance in terms of sequencing, ATCOs generally trusted its suggestions during the experiments.
- In terms of Human-AI Teaming, the nature of the interaction between the controller and IA showed that when ISA provided a suggestion, the controllers would decide whether to accept or reject it. This confirms that ISA must only operate to enhance decision-making, and shall never replace the actual ATCO in making decisions, in line with EASA’s Level 1B of Human-AI Teaming, which is *Human cognitive assistance in decision and action selection*.
- Even though ATCOs would generally trust ISA if its performance is on point, they would always need to have the last word.

HMI

Regarding the HMI, valuable feedback was collected that will lead to the improvement of its interaction design and the information shown on the HMI.

- Overall, the ATCOs understood almost all the components of the interface pretty easily, and they seemed satisfied with the level of information provided.
- Some ambiguity regarding CLT implementations: although they were viewed positively, they were not always interpreted correctly.

- Some controllers expressed doubts regarding the function of the CLT1 and CLT3 levels, as some information was presented in a way that they deemed unclear.
- CLT6 was probably the worst one as all the ATCOs could not understand its presence and usefulness at all.
- The sequencer settings were also welcomed favourably, though it appears now that they are too generic and need to be properly contextualised on Alicante's regulations.

6.7 UC#4 Validation 1 Conclusion

Upon completion of VAL1, these are the main lessons learned that point to new design directions:

Table 19. UC#4 lessons learned

Insight	Functional Requirements	Proposed solution for next iteration
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	Implementation of data pipeline and algorithms to process sequences in real-time.
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	On the HMI, ISA will only show numbers 1, 2 and 3 in the sequence.
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 tricky sequences and should not constantly attract ATCOs' attention when easy sequences are involved.	ISA will trigger alerts only during sequence switches in tricky situations, and not every time there's a sequence change.
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	ATCOs will be able to select "Reduced" and "Standard" modes for sequencing to be in line with Alicante's regulations. A button to switch ISA on and off will also be implemented on the HMI

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	CLT1 will only show the main KPIs (expected landings and arrivals) in the next 15-30-60 minutes
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	CLT3 will show essential data involved in the sequence change in a better layout, easier to consume
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	CLT5-6 will be redesigned or dropped completely, if no useful solution is found

For Validation 2, provided that all these solutions are implemented and ISA operates in real-time, the objective(s) will be to evaluate:

- Operational feasibility and acceptability more extensively, with real-time capabilities of the IA in scenarios where unforeseen events could change the sequence in every moment. For Validation 2, there will be scenarios that will dynamically change, forcing ISA (and the ATCOs) to adapt.
- Quantitative KPIs will have to be recorded thoroughly in VAL 2 during simulations with real-time ISA. The objective is to increase the number of operations (arrivals, departures), while reducing the number of safety events.
- To gather more info about safety by investigating tricky situations which could lead to safety events. The goal would be to understand if these safety events could be prevented consistently while using ISA.

To have participants interact directly with the HMI to gather more concrete evidence about the usability and user experience of the system.

7. Use Case #5 – Airport Safety Watch

7.1 UC#5 Validation 1 Plan Summary

Background: London Luton Airport & the Luton Safety Stack

London Luton Airport (LTN) is the fifth busiest airport in the UK, operating a single runway for commercial and business operations, and is a hub for low-cost airlines including EasyJet, TUI fly, Wizzair and Ryanair, alongside a number of business jet operators and ground handling services, with air traffic services provided by NATS.

Figure SEQ Figure * ARABIC 44. London Luton airport

London Luton Airport (the company: LLA) 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, and summarised on a weekly basis as part of its safety management practices, and recorded on a shared data platform called OPSCOM. Generally, analysis of this data is undertaken manually, and reviewed with airport Partners via key safety meetings such as the LTN Safety Stack, which meets quarterly to focus on safety management and safety improvement. The Stack, as it is known, led by LLA but supported as needed by other Stack Partners, has engaged with HAIKU to participate in this Use Case 5. The Stack is therefore the primary client for the output of UC5.

Research Problem: Enhancing safety learning and safety event predictability

Safety performance at LTN is good, but there are three incidents on the Stack's safety radar that are hard to eradicate. The first is **Hold-Point Bust**, wherein an aircraft is instructed to proceed to a hold-point on the taxiway system and await further instructions, but the hold-point continues through the hold-point. This has the potential to result in a taxiway collision or, in the worst case if the hold-point is just short of the runway, a runway collision, as appears to have happened this year in Tokyo's Haneda airport. The second is **turning the wrong way in the taxiway system**, which can also result in a collision or (far more likely) an aircraft having to take a circuitous route to where it should be, resulting in delays on the ground if the airport is busy. The third is **pushback error**, for example when an aircraft is pushed back by a tug from its gate, but is reversed to the left instead of the right, so that it ends up facing the wrong way on the taxiway system. These three incident types are illustrated below (Figure 45).

Figure 45. Accident types at LTN

LLA has experimented with different ways of visualising their safety data, including Safety Dashboards as also developed in the EU project Future Sky Safety. However, manual analysis of the data, much of which represents 'weak signals' since such events themselves are rare, has not provided effective safety interventions.

The overall aim of UC5, therefore, is to develop an **Airport Safety Watch** (ASW) to assist the airport safety duty-holder (LLA) and other principal airport users (the Stack Partners), in reducing the risks of these three key incident types (pushback error, hold-point busts and incorrect taxiway selection). ASW aims to do this in two phases: first, to assist in deeper data analysis using data science and visualisation techniques to determine causal and contributory factor patterns, leading to actionable safety improvement measures that will reduce risk. This first phase is the subject of VAL1. The second phase of ASW development is to develop a predictive warning capability that could be used in real-time or on a daily planning basis, identifying when there is a heightened risk of any of these three occurrences. Phase 2 will be the subject of VAL2 later in 2024.

Key R&D Objectives & Requirements

The VAL1 Plan is informed by three sets of considerations: R&D objectives, validation requirements, and performance measures and metrics. Key R&D Objectives as defined in D6.1 are show in the table below (Table 20), with objectives 01 and 03 relating to VAL1:

Table 20. UC#5 validation objectives

OBJ-ID	Validation Objective	Success Criteria ID	Success Criteria
UC5-OBJ-01	To assess the Operational Feasibility and Acceptability of the solution from the Airport perspective. An underlying R&D objective here is to see how, why and to what extent collaborative entities (companies) accept AI-derived safety intelligence.	UC1-CTR-01	A sufficient quorum of Stack members agree that the insight / solution has merit and are willing to explore its further exploration and/or implementation.
UC5-OBJ-02	To assess the ability of the system to 'see around the corner' and predict new events or when existing event types will have a higher likelihood of occurrence.	UC1-CTR-02	Stack partners take the intel seriously enough to increase monitoring during higher-risk periods and issue warnings to operational personnel, and/or consider operational mitigations during high-risk periods. Additionally, this results in lower incident rates.
UC5-OBJ-03	To assess the degree of new insight afforded by the AI support. The R&D objective relates directly to HAT, in that the AI may point out new ways of understanding the data, while the operational players interpret this and derive realistic solutions. The solution may therefore be truly dependent on Human AI Teaming, since neither party can fully solve the problem alone.	UC1-CTR-03	The LLA Safety Team and related safety personnel in the Stack (e.g. airlines, NATS, Ground Handlers) agree the solution is novel or is framed in a new way not previously considered.
UC5-OBJ-04	To see the degree to which the HAT interactions lead to new safety learning avenues, via changes to reporting and recording systems.	UC1-CTR-04	The interactions and analytic iterations on the 3 incident types lead to changes in reporting and recording systems, with a broader set of factors (e.g. on traffic parameters) or a higher granularity of factors (e.g. on human performance aspects).

The objectives have been broken down into requirements in D6.1, and the two related to VAL1 are shown below:

Table 21. UC#5 Requirements

HL-REQ-ID	**UC5-HLR-01**
Requirement	Sufficient data are acquired from LLA & Airport Stack Partners to share with the technical partners (ENG/SUITE5)
Rationale	For the airport safety watch concept to be viable, it needs sufficient quality data on all incident occurrences and situational data. In practice, this involves sharing data, at least of the last 7 years, such as pushback errors, selection of wrong taxiway and holdpoint busts, and additional data relating to the events as required.
KPA	Operational, Technical, Human performance, Safety. E.g. traffic movements, meteo, time of day, etc., Safety (near misses and incidents/accidents), Human Performance (errors, recoveries; time-on-shift; roster information (location in shift cycle); operational role (driver; flight crew, controller etc.)
HL-REQ-ID	**UC5-HLR-02**
Requirement	Identification of new solutions for incidents: pushback errors, selection of wrong taxiway, holding-point busts.
Rationale	Identifying new solutions or investigation of new avenues to decrease their occurrence rates of the incidents
KPA	Safety, Human performance, Operational

The performance metrics identified in D6.1 are shown below:

Table 22. UC#5 performance metrics

KPA	**Category**	**KPI**
Safety Performance	MoP	Incident rate per quarter, for the following three types: Incorrect taxiway selection Holding-point bust Incorrect pushback
Safety Management	MoE/MoP	Concrete plans for incident rate reduction. Action Implementation Timescale
Safety Management	MoE/MoP	Changes to reporting practices to include new incident contributory factors
Stack Safety Collaboration	MoE	Engagement by the Stack on the insights from the tool, and collaborative efforts on risk reduction
Human-AI Teaming	MoE	Degree to which human and AI can collaborate to extract insights and warnings
Human-AI Teaming	MoP	Degree to which AI-produced warnings are heeded by operational partners in time to realise a risk reduction/mitigation.

Summary of UC5 VAL1 Plan

The VAL1 plan can be summarised as follows:

Table 23. UC#5 VAL1 plan summary

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
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
4	Increase in Human-AI Teaming	Development of preliminary Dashboards and Interactive Support media.	Partners ask for tools they could use in the future, and begin to specify requirements and use cases.

7.2 Deviation from Initial Validation Plan

The only deviation from the original VAL1 relates to R&D Objective 04, to see if reporting and investigation processes can be updated with new requirements on key contributory factors to allow better ‘forward vision’ in the future. This is seen as premature at the moment, and so has been pushed back to VAL2.

7.3 UC#5 Demonstrator and Assistant Features

The two figures below show examples of the UC5 ‘Demonstrator’ under development at the moment. The first figure shows an Interactive Dashboard, wherein incident types can be examined from a number of perspectives and via different visualisations. The second figure shows an Interactive Map Display which is being developed to ‘zero in’ on particular incidents and also geographical ‘hotspots’ on the airport surface (Apron, Taxiways and Stands). Both of these displays are still under development at the moment, pending the next Stack meeting to firm up on requirements (how they wish to see and explore the different data visualisations) from the Stack Partners. This interactive map, when implemented, will be able to show the origin (stand) and route of an aircraft that was subsequently involved in an incident, and whether it was the normal route or was different, e.g due to construction or other factors.

Figure 46. UC#5 demonstrator interactive map display

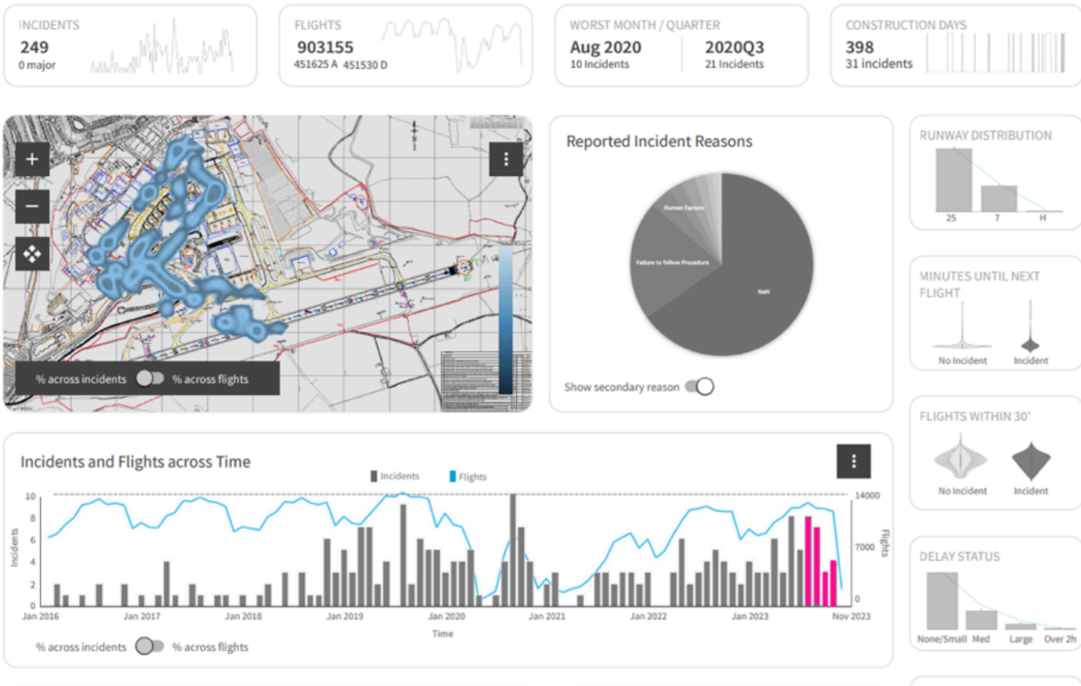


Figure 47. UC#5 demonstrator interactive dashboard

7.4 UC#5 TRL

As far as TRL evolution is concerned, by the end of March 2024 ASW has reached TRL5. It reached TRL3 (Proof of Concept) in the meeting of July 18th 2023, and TRL4 (Functional Verification) by the Stack a subsequent meeting with LLA and NATS in November 2023 to consider functional adequacy, and by March 2024 reached TRL5 (breadboards verification [by LLA] of reduced scale in a relevant operational environment).

7.5 UC#5 Validation 1 Activities

The Table below gives an overview of the activities and the progress made so far:

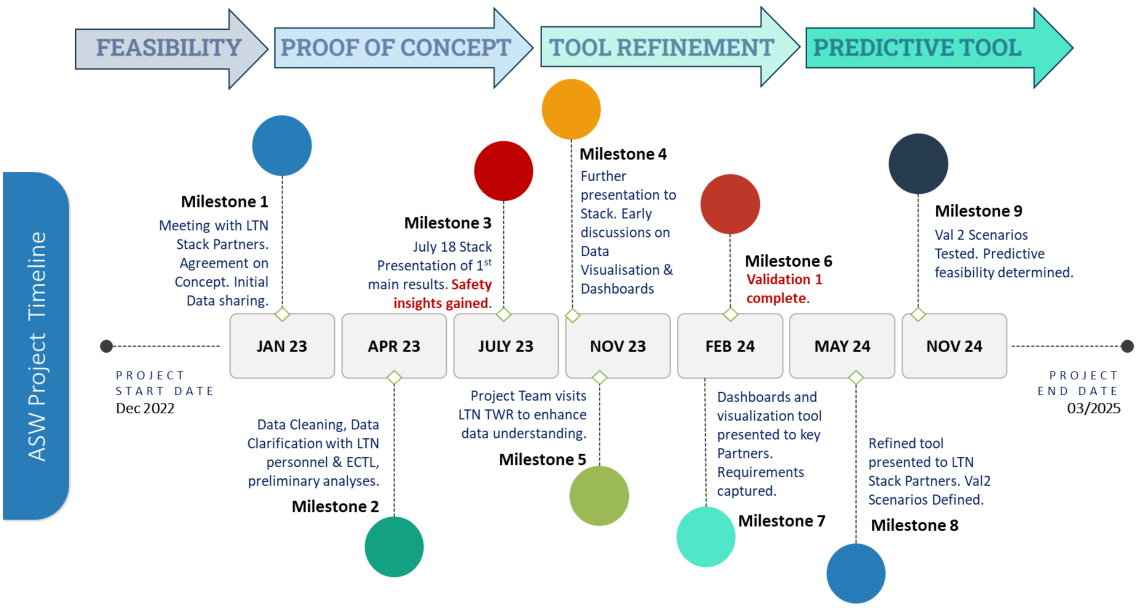


Figure 48. UC#5 activities overview

Since the Stack Meetings are every quarter, this creates a certain ‘rhythm’ for the Use Case, wherein there is an intense period of interaction (often for just one day) followed by detailed offline analysis and development interspersed with clarification queries from the Project Team to LLA, for example. The Stack meetings themselves are always ‘Presential’ (in-person), though sometimes SUITE5/ENG join remotely (ECTL co-chairs the Stack so is always present at the meetings). The July 18th 2023 Stack Meeting involved all Project parties being present, and was a ‘watershed’ moment for the project, wherein SUITE5 and ENG presented their data visualisations and interpretations of the data, and the Stack Partners ‘met them halfway’ via their own operational interpretations, leading to safety improvement insights. Another Stack meeting in November was also highly useful for the Use Case, especially as it allowed the Project parties to visit the LTN Tower and observe how the various operations occurred in real time, and afforded time to ask tower controllers detailed questions as well as ‘listen in’ to Tower-aircraft voice communications, instructions and readbacks.

Additionally there have been numerous online meetings between the three project parties (ENG, SUITE5 and ECTL).

7.6 UC#5 Validation 1 Results

The validation results are documented below for each of the 4 VAL1 elements.

Stakeholder Engagement

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

HAIKU was first presented at the Stack's 24th meeting in Luton on Jan 26th, with 15 Stack Partners present at the meeting. There was general interest from a number of Partners, and agreement on the selection of the three incident types to be studied. The next discussion of HAIKU was two meetings later on July 18th, when the first results were presented. This presentation and discussion gained considerable traction in the meeting, with active engagement from NATS and airlines and the business jet community, with some of the Ground Handlers saying they hoped to be involved at a later stage if possible.


24th LUTON SAFETY STACK MEETING
26th January 2023
Putteridge Bury Conference Centre
Hitchin Rd, Luton LU2 8LE

AGENDA

09:00	*Tea/Coffee*
09:30	Welcome & News (Nick/Barry)
09:45	Review of previous minutes and actions (Nick)
10:00	Dublin Airport Safety Stack report (Barry)
10:15	Safety Performance Update and 2022 review (Liam/Alistair)
10:45	*Coffee*
11:00	Review of Action Lines (Nick) 4 – Broken Windows 9 – Safety Leaders Programme
11:15	Impairment Policy Review (Nick/Liam)
11:45	Can Artificial Intelligence raise the Stack's game in safety? (Barry & SUITE-5)
12:30	*Lunch*

Figure 49. 24th Luton Safety Stack Meeting program

HAIKU was also presented at the following Stack meeting in November, and is scheduled to be discussed at the meeting on 27 February 2024. Several Partners have supplied data to the Use Case, especially LLA, but also NATS, and airlines have offered data as needed. NATS also invited SUITE5 and ENG to visit the Tower to see how the operations were carried out, and accordingly a half-day visit took place on 7th November 2023.

Data Sufficiency

VAL1 Element	Description	Activity	Success Measures
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

LLA decided to offer data from 2016, as that was the first year where there had been an increase in the quality and trustworthiness of the data. Therefore, 7 years of data were available. Much of the data was transmitted in excel format, and there was considerable need for data cleaning and preparation for data science analysis as shown in the figure below (Figure 50).

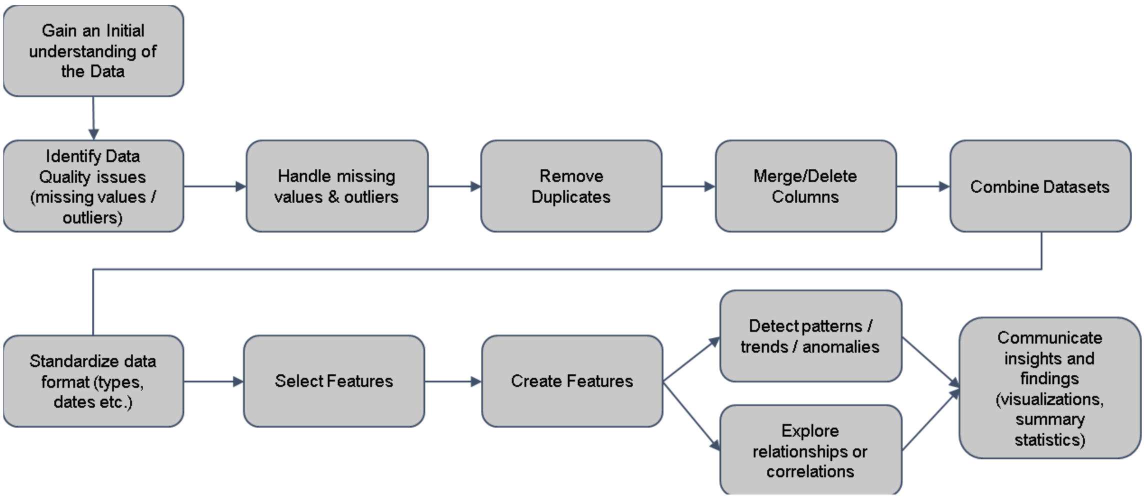


Figure 50. UC#5 data preparation process

Once the data were prepared, analysis could begin, which included many exchanges between the data scientists and LLA to correctly interpret what the different data sources meant and to see how they could be correlated. Some examples of how the evolving data analyses looked are shown in Figure 51; these and many other visualisations of the data were presented at the Stack meeting on 18th January.

Overall, more data would be desirable, particularly if there were more incidents (thankfully they are rare). This was a learning point for the non-data scientists, in that a large amount of data doesn't help if events are really rare, since forging correlations becomes difficult. On the other side, sometimes data scientists might request data that for operational people had no safety relevance, e.g. number of passengers on aircraft. Occasionally, e.g. searching for meteorological data, this could be gained from open source sites.

On balance, there has been sufficient data to prove the concept of ASW in terms of gaining new insights, but there may be insufficient (at this moment) to develop a predictive tool (a VAL2 objective).Nevertheless, new data continues to be added into ASW, with the latest data upload bringing all incidents until end of June 2024 into the ASW data 'pipeline', feeding directly into the latest version of the Dashboard..

New Safety Insights

VAL1 Element	Description	Activity	Success Measures
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

Figure 51. UC#5 sample of evolving data analysis

The Stack meeting on July 18 2023 was particularly constructive and generative, leading to the identification of several new safety insights, as detailed in Table 24.

Table 24. New safety insights after the Stack meeting

Issue	Discussion	Partners	Related Requirement and/or Validation Objective
Incorrect Taxiway Selection	It was remarked that although LTN is not a highly complex airport with multiple runways etc., it does have a relatively high number of junctions, which can perhaps lead to confusion or perception errors about where aircraft believe they are and where they should go next. In some larger international airports they operate a 'follow the green' system, though it is not clear that LTN could adopt such a system. In the future however, the airport will gain an ASMGCS (Airport Surface Movement Ground Control System) which gives a live-updated map of the airport surface and all aircraft (and some vehicles). This may also be augmented by CCTV particularly around 'hotspots' and to those areas that are difficult to see from the Tower.	Airlines, ATC, LLA	UC5-HLR-01 UC5-HLR-02 UC5-OBJ-01 UC5-OBJ-03
Hold-point Bust	There was some discussion of hold-point busts and practices at other airports. A number of European airports these days have 'zones' which the aircraft crosses into and where it waits, rather than a line that the aircraft should not cross. Pilots more familiar with these zones or areas may inadvertently cross over a holding point, thinking they are supposed to enter a zone.	Airlines, ATC, LLA	UC5-HLR-01 UC5-HLR-02 UC5-OBJ-01 UC5-OBJ-03
Pushback Error	Stands 62 and 71 were highlighted by the data analysis presentation as being more prone to pushback error. Partners noted that Stand 62 has no sign, which might contribute to error rates (it is for business jets rather than commercial jets, and many business jet pilots are unfamiliar with LTN's layout etc.). Stand 71 is at the end (a cul-de-sac) and also might not be as well signposted as other stands. For pushback error, it was also noted that stands that occur on a bend can be tricky. In some airports the pilots no longer control the direction in which they are pushed back, and it is left to the ground handlers. One or two Stack partners could see the advantage of this, as the local staff are more familiar, and there can be misunderstanding when communicating with flight crew about what is left or right. The best form of instruction was also discussed, as to whether it should be 'left' or 'right', or a compass reference (e.g. East, South, etc.). One	Airlines, ATC, LLA, Business Jets	UC5-HLR-01 UC5-HLR-02 UC5-OBJ-01 UC5-OBJ-03

	further suggestion was to pushback to a landmark, which could be a clearer and less confusing form of instruction.		

Since the July meeting, signage has been improved at the two Stands mentioned, as well as at one key point on the taxiway system close to the runway, as it was recognised that it was not clear to all pilots (many of whom fly to LTN infrequently). For Hold-Point Busts, this is still a work in progress to understand the factors that lead to this event, and so each new event is added to the database as it is released.

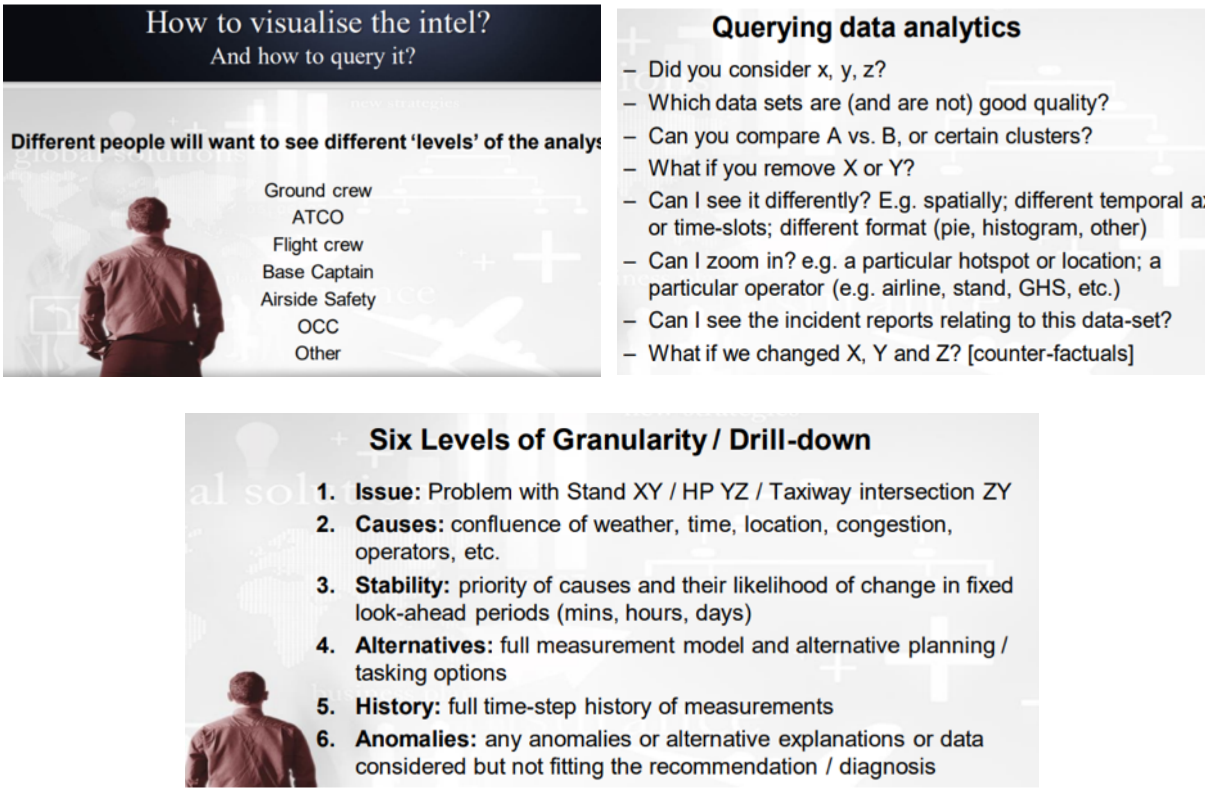
Some interesting factors raised by the analysis are still being explored, e.g. there appear to be more events when the runway is being used in the less frequent direction (since aircraft take-off and land facing the wind, if the prevailing wind shifts, the direction of take-off/landing may be switched 180 degrees. Operation in this less frequent mode may be more productive of incidents.

One key factor related to both incorrect taxiway selection and hold-point bust, is the route the aircraft took prior to the event, i.e. from the stand to the event point on the taxiway. This is not currently recorded by incident investigators, but it could be a key element. This is being discussed with LLA/NATS, and recording of such data in the future via an interactive map the Use Case has developed, may enhance both forensic and predictive analysis in the future. At the moment, this is the only potential 'new data recording' suggestion arising from the study (a Val2 objective), to enhance safety learning.

VAL1 Element	Description	Activity	Success Measures
4	Increase in Human-AI Teaming	Development of preliminary Dashboards and Interactive Support media.	Partners ask for tools they could use in the future, and begin to specify requirements.

At the outset of UC5, it was not clear whether there would be any HAT at all in UC5, as one option was simply that ASW would be a black box that would deliver insights (phase 1) and predictive warnings (phase 2). However, what has become clear, at least for the Stack Partners, is that they wish to see and interrogate the data derived and displayed. Indeed, the insights gained have come from a human-AI partnership, with operational people surveying the data science-derived visualisations of data, rather than passively waiting for 'magic answers'. This has led to the request for an Interactive Dashboard, which is being developed now by ENG and SUITE5, and is to be presented at the next Stack meeting on Feb 27.

What remains to evolve is exactly who uses it, and what information and data interrogation capabilities they find most useful. In terms of who will be the users, at the moment it is primarily LLA, which makes sense as they are the duty holder for safety, but it could be of use for airlines and the ATC provider in particular. For such organisations, there are then a further set of potential users, e.g. for an airline it could be the Base Captain, Ground Ops Safety, the Safety Director, or someone in the Operations Control Centre (OCC). For such different users, they may wish to interrogate the data using questions such as those in the second box in Figure 52. More technical users (e.g. safety departments) may also be interested in different levels of explainability of the data, as shown in Figure 52 (based on Construal Level Theory). As ASW develops, requirements will be gained in these areas (who wants to use it; how do they want to interrogate/manipulate/visualise the data; how might they need to verify it / explore data model quality and alternative 'explanations'). The more ASW is developed along such lines, the more serious it is being taken by operational organisations, and the more likely it is to have a real impact on operational safety.



At the end of Val1, informal feedback was obtained from one airline (WIZZAIR), the ATC provider (NATS) and LLA, the airport safety duty-holder. What became clear was that ASW should be a tool primarily for LLA. Some of the other Stack Partners (especially airlines and ATC) already have their own individual bespoke legacy incident reporting and learning systems, whereas others (the Ground Handlers) often have less developed systems. It is only LLA who gains the overall picture, and has the specific duty to look across companies to learn lessons. Other Stack members are therefore largely content to see the overall picture and learn from it where possible, but to let LLA take the lead in using and analysing the ASW Dashboard. This feedback led to a certain clarity in the project, and allowed the Dashboard to develop more rapidly following Val 1, to a much more mature and tailored system for LLA, to be evaluated in Val2.

This 'course change' and focus on the airside safety manager, also renders the tool less grounded in LTN airport itself, and more amenable to uptake by other airports. In the

February 2024 Stack meeting, Dublin Airport (who also have a Safety Stack) sent a delegation to Luton, and were impressed by the ASW Dashboard and are considering whether they should embark on a similar venture. They are waiting to see the final version. As ASW matures it will be shown to other airports via the EUROCONTROL Safety Team (safety directors and safety managers from all the Air Navigation Service Providers in Europe). Additionally, there is now a recent ICAO action to develop and make available Safety Stack promotional guidance material for all airports, and the ASW approach will be referred to in this material (initial draft due October 15 2024).

7.7 UC#5 Validation 1 Conclusion

VAl1 results serve as a proof of concept for phase 1 of the ASW research concept: sufficient data could be analysed in novel ways to allow operational users to identify new safety solutions to existing, long-standing safety issues. In particular, the results have led to strong stakeholder engagement. The challenge now is to ensure that such engagement is maintained. This can be realised via delivery of usable Dashboards and associated media (e.g. interactive map) which allows further safety insights to be identified. The design of such Dashboards needs to be user-centric, and so requirements need to be gathered from interested users. It may be useful to further develop such requirements in the context of key incident types, in particular hold-point busts and incorrect taxiway selection, since pushback error already has a 'win' via the safety insights generated by the analysis.

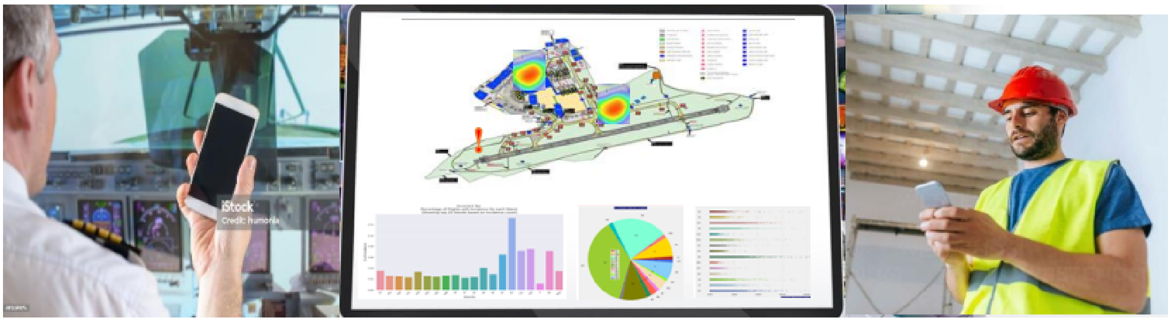


Figure 53. User-centred design

In terms of Phase 2, evolving ASW towards becoming a predictive tool, this is a major step beyond where ASW is now, and would require real-time data feed of 'clean' data from multiple partners. It may be that an intermediary step is more realistic, of using ASW to identify clusters of factors that would be high risk. Such characterisation of factors could then be used to 'de-risk' operations on days when such factors are likely to coincide, including scheduled maintenance on the apron, taxiway or stands. If such intermediary approaches could be found to work and be seen as useful, this would still be predictive use, and might pave the way for one day having a real-time predictive function.

In terms of HAT, what is needed is working on the requirements of those who are most seriously thinking about using it, namely LLA, with the ATC service provider being the next most interested party.

The ASW could also be useful for investigation following incidents, both from the perspective of giving a more systemic understanding of what happened and the factors apparent, and also from a more practical viewpoint in terms of the interactive map display for taxiway and hold point-related incidents, since it can easily create a clear visualisation of the path taken, which over time will show not only where the hotspots are, but the pathways most likely to lead to incidents.

In summary, ASW has shown that it can lead to new safety insights, through a HAT relationship with key operational personnel. The next step is to refine the interactive tools to optimise the HAT, and move towards a predictive concept. Additional use of ASW for investigation may be explored, as well as refining the data recording needs of ASW in the future.

The findings from UC#5 VAL1 are summarised in the table below.

Table 25. UC#5 lessons learned

Insight	Functional Requirements	Proposed solution for next iteration
The ASW research concept has led to strong stakeholders engagement with safety.	Maintain the stakeholder engagement with safety.	Design of usable dashboards and associated media (e.g. interactive map) which allows further safety insights to be identified.
ASW should become a predictive tool.	Increase safety predictability with ASW.	Characterise factors that could be used to 'de-risk' operations on days when such factors are likely to coincide, including scheduled maintenance on the apron, taxiway or stands.
ASW could also be useful for investigation following incidents.	Create visualisation for investigating incidents.	Use the interactive map display for taxiway and hold point-related incidents, since it can easily create a clear visualisation of the path taken, which over time will show not only where the hotspots are, but the pathways most likely to lead to incidents.

8. Use Case #6 – Airport Spreading Virus Prevention

8.1 Overview of Validation Activities

The COVAID tool as presented in UC6 is expected to reach a TRL5/6 at the end of the project. It started as a novel application with TRL0 and it was expected to reach TRL2 after the development and the VAL1. To this end we aimed to test the various components of our UC in a confined laboratory environment for VAL1 in order to validate our components. Note that verification and validation are two stages that need to exist in software product development. Verification is "Are we building the product right?" meaning if the software does not have any bugs and if it produces results. This has been documented in D4.4. Validation is "Are we building the right product? Meaning that all the individual components need to be validated to output the desired results. An example of the validation for UC6 is if the weighting factor produces the right results, starting from the input from the passenger and the sensors.

The validation activities of VAL1 for the UC6 mainly focused on the validation of components, since they have to be verified and validated before being shown to the users. The diversity of UC6 tasks make it a tough integrating process which needs to be finalised before being accessible to the students/candidate pilots of the Amygdaleona airport in Kavala. In the validation we included validation results of the components as well as some kind of verification for the reviewer to have a clear view of the UC6 tasks. The application is set to be ready for the VAL2 and will include a number of test passengers. The air quality tool will be given to the staff of the airport for validation purposes, even though we need to have air quality domain experts to determine the efficacy of the output of the tool. In the figure below you will see the validation plans with respect to the WPs 4 and 6. The complete timeline for VAL2 will be given in the validation plan that will be refined for VAL2.

Intermediate steps between VAL1 and VAL2 may include the availability of the tools to the employees of the Hellenic Institute of Transport to give us feedback regarding the validated components of our application. Actions that will take place as well are the deployment of the solution to the Amygdaleona airport and the production of two wireless networks one with the cameras and the other with the air quality sensorboards.

Stakeholders that include the passengers (student/pilots) as well as the staff of the airport will have access to the application and we will provide a survey to identify trends that may emerge with the use of the tools. Other stakeholders will not be included for this UC in order to keep it simple. For future validations we will attempt to use a larger scale airport, and the stakeholders will include store assistants and owners as well as staff of the airport.

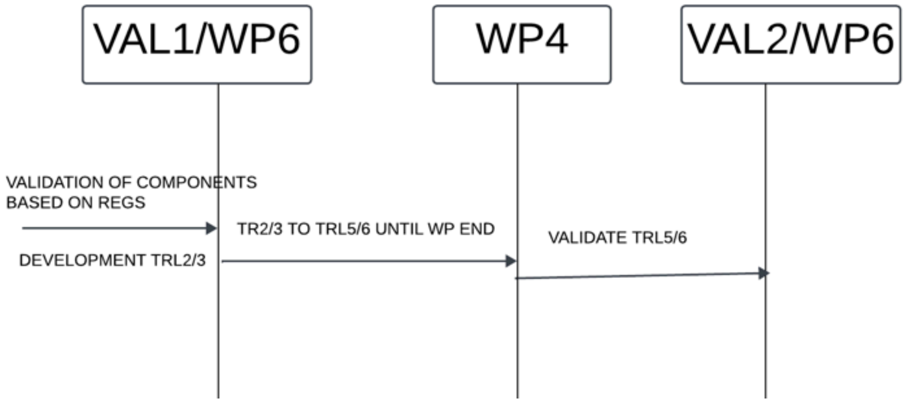


Figure 54. Validation phases and TRL.

Regarding the validation activities of our UC the contributions are highlighted below:

- **Hardware Components Validation** which comprises the cameras validation and the indoor air quality sensorboard validation

- **Software Components Validation** which consists of the following:
  - ML Infection Classifier Validation
  - ML Air Quality Classifier Validation
  - Lidar Validation Activity
  - Weighting Factor Validation
  - Web Service Validation

- **System Validation** which comprises the Android Application Validation and the chatbot communication validation as well as the air quality statistical, classification and explainability tool validation.

The VAL1 of the COVAID toolset has been done taking the following KPAs in consideration.

Table 26. UC#6 performance targets

KPA	Category	KPI
Management Efficiency	MoE	Number of people routes to each common space and reduction of crowded spaces as given by out weighting factor
Safety	MoE	Reduction of overcrowded spaces by the recommendations of the COVAID tool
Human Performance	MoP	Passenger trust of the routing tool Optimal selection of common places visit according to the tools AI recommendation Usability due to its simplicity Health and safety operator to check the state of the air quality in indoor spaces and explanations given by the tool

8.2 UC#6 TRL

The components that have been characterised by TRL for UC6 are given below:

- HMI for android application and air quality tool.
- Routing algorithm.
- XAI using SHAP
- XAI using CLT
- Wireless Camera and air quality prototypes.
- ML algorithms for classification.

In terms of the HMI for the Android application the current TRL level is 2. Even though the tabs have been successfully implemented and the routing algorithm has been successfully interfaced from the cloud to the application, as well as the chatbot exhibited functionality the readiness of the application is low. This is due to the fact that it has been tested with a small number of staff members in the lab and that it will definitely require modifications. For instance, the routing algorithm may need to provide more data to the application to be transferred to the chatbot for CLT explainability. Moreover, the database that is used will be substituted with a more concise one with many more entries and it may be in the form of a CSV file that will be assessed for the best answer to the passenger's comment or question. The air quality tool will take input from the devices that will be placed at the Amygdaleona airport in Kavala Greece and more explainability features will be added, even though it is a standalone application that the data will arrive as plug-n-play.

The routing algorithm is again in TRL 2 since it has been proven to work in simulation and with initial experiments by simulating passenger traffic and occupancy. VAL 1 showed that the algorithm works in a controlled setting and the higher level of TRL will be ensured with the deployment at the aforementioned airport.

The XAI using the SHAP method is at TRL 3. This is the case since the SHAP method is established and code as well as documentation is available online. In terms of the data fed to the SHAP fragments of code, the method showed the significance of the features of the dataset and it was shown that it was satisfactory. The fact that data coming from the devices built will feed the method to show the results constitute the application to TRL 3. The deployment at the airport will ensure higher TRL.

The XAI using CLT levels was at a quite low TRL (1), since only the CLT 1 was implemented successfully. To this day, the CLT levels appear at the chatbot providing the relevant explanations to the passenger and high TRL (5) will be ensured. The database of the chatbot and the incoming data suggested this low TRL; The validation of the data sources and output have been validated but the level of presentation suggested a low TRL.

The wireless camera and indoor air quality prototypes are at TRL 3. The main reason is that the devices consist of off-the-shelf components that have been validated without required oscilloscopes and spectrum analysers. The wireless camera prototype comprises a mini PC, a commercial camera and a wireless USB stick. These do not require special validation. In terms of the software built, the tracking exhibited a problem in live feed where each frame showed a different ID. To this day, we built a new algorithm which we are validating in live feed by using smart tracking without a well-known tracker. These developments constitute the wireless camera system as TRL 3. The indoor air quality sensor also comprises commercial components. The respective software has been validated with other devices and measurements that showed a proper functionality. However, the deployment of the sensorboard in conjunction with the networking software that is being built will raise the TRL.

The ML algorithms for both the COVID likelihood and the air quality classification are at TRL 3. This is due to the fact that the development of these algorithms do not exhibit unknown functionalities due to novelty, since they are established as classifiers in the scientific community. The TRL3 is given since for the likelihood of infection the data used is artificial and for the indoor air quality a commercial dataset is utilised. Hence, it is wise to allow a rework plan when these algorithms will be operating in practice upon deployment of the entire system.

8.3 UC#6 Validation of Individual Components

In this section we state the validation activities of the developed and selected hardware and software components of this UC.

Hardware Components Validation

In this subsection we provide the reviewer with the hardware components validation of our UC.

Cameras Validation

In order to test the wireless camera prototype, we had to separate the tasks into two subtasks. The first one was the object detection and tracking and the second was the wireless communication. For the former, we had to devise an algorithm that would count the persons in the view of the camera in a unique manner. In order to do that, we obtained a video that showed people walking in both directions (left to right). The algorithm we devised is a very simple one with the x1 and x2 of the rounding box of the object having to be between a line that exists in both directions as it can be seen in Figure 55.

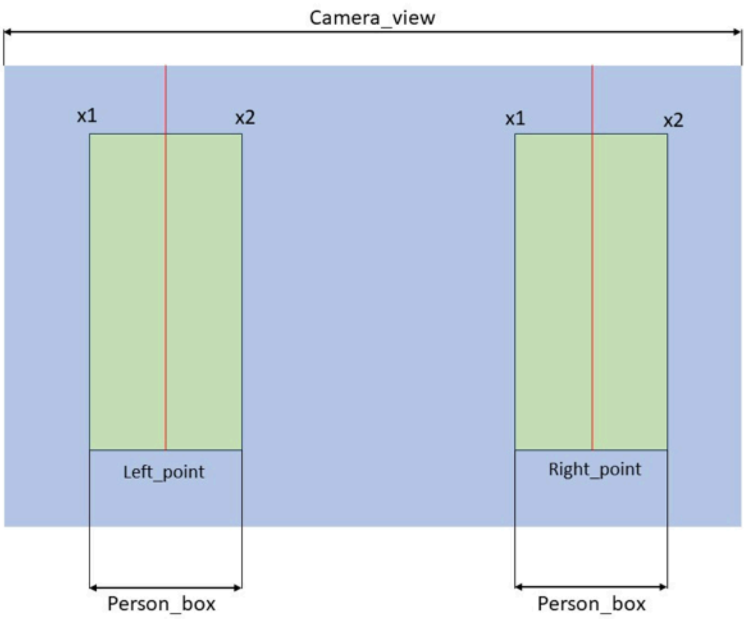


Figure 55. Bounds of vision algorithm

As the reader can see from the Figure there are two lines depicted as left and right point respectively. When the person's rounding box is between these two lines, the algorithm puts the ID from the DeeOCSort tracker to a list. Note that for each frame the list is checked not to allow duplicates. In the case that the person is moving towards the left, if the x1 point of the rounding box crosses the line in terms of pixels and the x2 does not again in terms of pixels the counter is incremented by 1. In a similar manner, the algorithm checks these conditions for the right side. It has to be pointed out that there is a third condition to be satisfied for the successful counting of persons, which is for the ID to reside in the list. Then the ID is removed from the list and the person is not counted twice. For the case that the person is moving from left to right and the points x1 and x2 areas they can be seen in the Figure, the person is not counted since the ID is not in the list because the rounding box has not been between the left and right points lines. In the same manner for the opposite direction. A demonstration video[3] can be seen at https://shorturl.at/blFST.

During our experimentation with streaming using a web camera, a significant challenge arose as the tracker id changed for each person in consecutive frames. This hindered the efficiency of our approach, prompting us to seek an alternative unique identifier. Our solution involved counting the number of detections, with each detection being assigned a unique id in ascending order based on the individual's presence on the screen. Despite this, the presence of multiple individuals led to fluctuating IDs. Our preliminary attempts showed no improvement to the maintenance of the IDs. It is highly probable that the camera does not maintain the ID and changes it per frame.

[3] The original material can be seen on youtube at http://youtube.com/watch?v=VamCnoHZezg

In our most recent approach, we addressed these challenges by implementing an object detection solution that incorporates a Single Shot Detector (SSD) using the MobileNet architecture. This technique excels in detecting objects in a single pass, requiring only one step for generating region proposals and another for object detection within each proposal. Compared to two-shot detectors like R-CNN, SSD offers notable speed advantages.

To facilitate object tracking, we utilised the Centroid tracker, renowned for its reliability. Essentially, the centroid tracker calculates the centroid of bounding boxes, representing the (x, y) coordinates of objects in an image. Once our SSD provides these coordinates, the tracker computes the centroid (centre) of each bounding box, determining the centre of each detected object. Subsequently, a unique ID is assigned to each identified object, enabling seamless tracking across a sequence of frames.

We tested that with the same video and found that there was an extra detection of a person. This indicates that there is improvement to be done in this case, even though the tracker seems to be more reliable. Note that we had to make the code work horizontally rather than having a top view due to the low ceilings in our premises. If we have a setting with higher ceilings, we will provide a solution from a top view and a horizontal line as a detector. The reader can see a snapshot of the video detection in Figure 56.

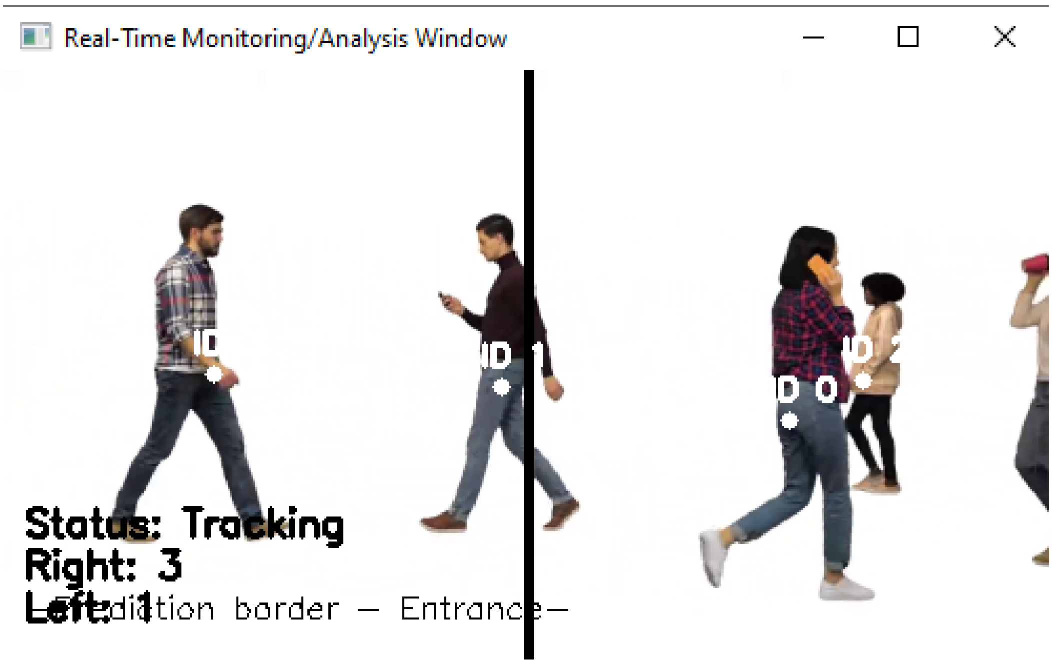


Figure 56. Image of video of real time monitoring

Initially we installed the cameras from V-Count (www.v-count.com) as the Figure 57 depicts. The company offers an API which can be utilised in order to obtain the real-time occupancy of the place it monitors. We installed three cameras but we set up the two cameras as inbound and outbound, hence we had two sites.

The cameras were validated for their accuracy by having persons going inbound and outbound, both single and in pairs. We noticed that there were calibration issues that emerged. One of the hallway cameras did not measure the person if she passed from the edge of the hallway and occasionally it counted three persons when two persons were walking side by side. Moreover, the room camera had the same problem when someone entered sideways. Finally, the other hallway camera counted people walking in parallel from the detected region of interest. These were calibration issues that were tackled with the assistance of the provider. We can see one of the problems that emerged in Figure 58 and Figure 59 as shown in the respective section of deliverable 4.4. Note that the red lines indicate the detection region and the yellow lines the problematic cases.

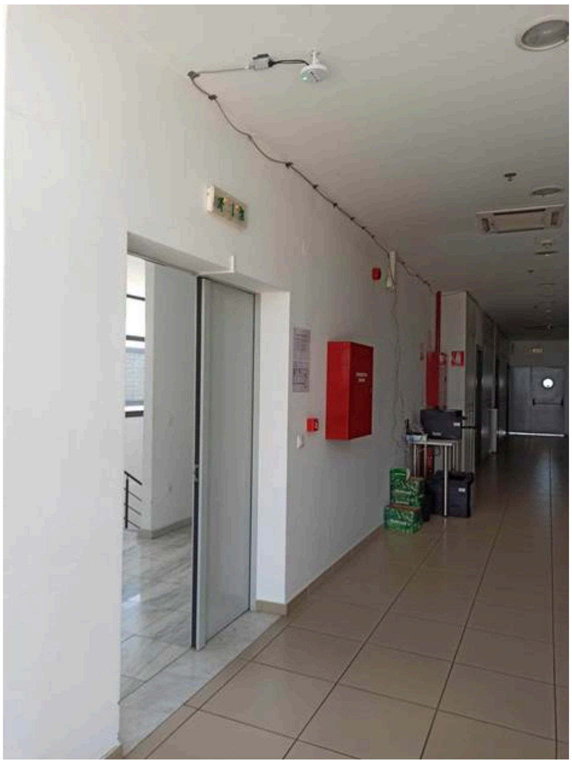


Figure 57. VCount person camera deployment

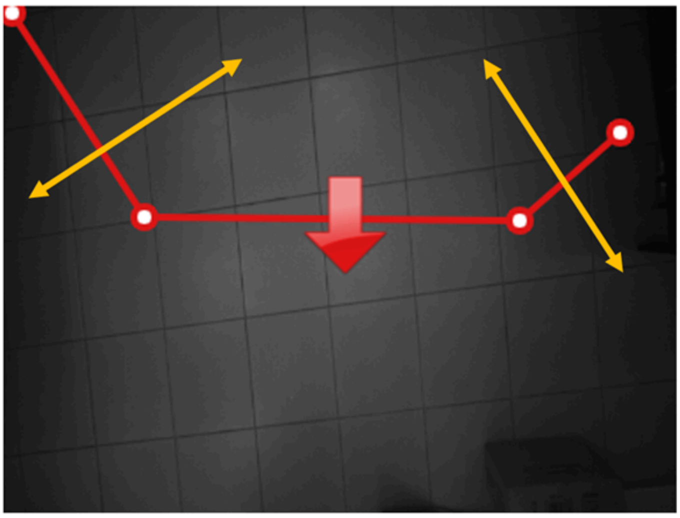


Figure 58. Room camera sideways passing detection problem

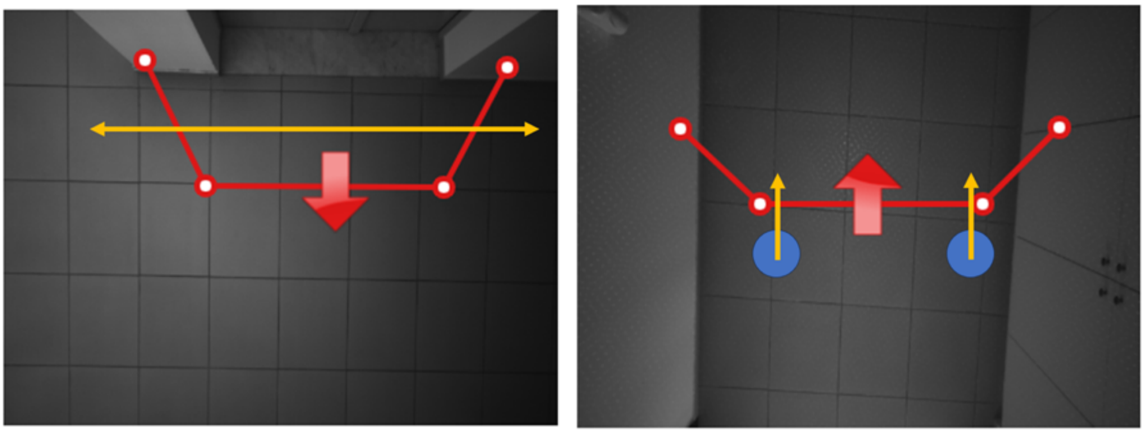


Figure 59. Hallway camera parallel passing detection (left) and two persons (right) passing problems

The API provided by VCount was quite straightforward and we managed to obtain the data we required. Below you can find the code for the retrieval of the occupancy of both sites and the writing to a csv file for further processing.

```
import requests
import json
import time
import csv

url1 = 'https://cloud.v-count.com/api/v4/realtime_occupancy_sec'

params1 = {"username" : "Certh", "password": "password", "store": "Site 1" }
params3 = {"username" : "Certh", "password": "password", "store": "site 3" }

f = open('test_vcount_cameras.csv', 'w')
writer = csv.writer(f)

while 1:
response1 = requests.post(url1, data=params1)
response3 = requests.post(url1, data=params3)

data1 = response1.json()
data3 = response3.json()
data_final = str(data1['occupancy']), str(data3['occupancy'])
writer.writerow(data_final)
time.sleep(60)

```

It is a simple code script to obtain the values we need using the API call with the POST method, which returns a JSON file and is processed accordingly. We cross-checked the data obtained from the script with the persons counting from the dashboard to validate its optimal operation.

The development with the calibration problems of the VCount cameras led us to purchase a different set from FootfallCam (www.footfallcam.com). Two of the cameras were installed in the same places as the previous one and the last one was installed further up the hallway and not close to a door, in order to mimic three sites.

The cameras use a different region of interest which is depicted in Figure 60 left. As it can be seen it covers the entire hallway and it tracks the person to a rectangular as opposed to the previous solution. In terms of the room camera the region is still greater offering a better detection as it can be depicted in Figure 60 right. We can see that the region of interest is covering the whole door and beyond with the rectangle that is shown in the Figure. It has to be stated that the two solutions that have been tested are GDPR compliant.

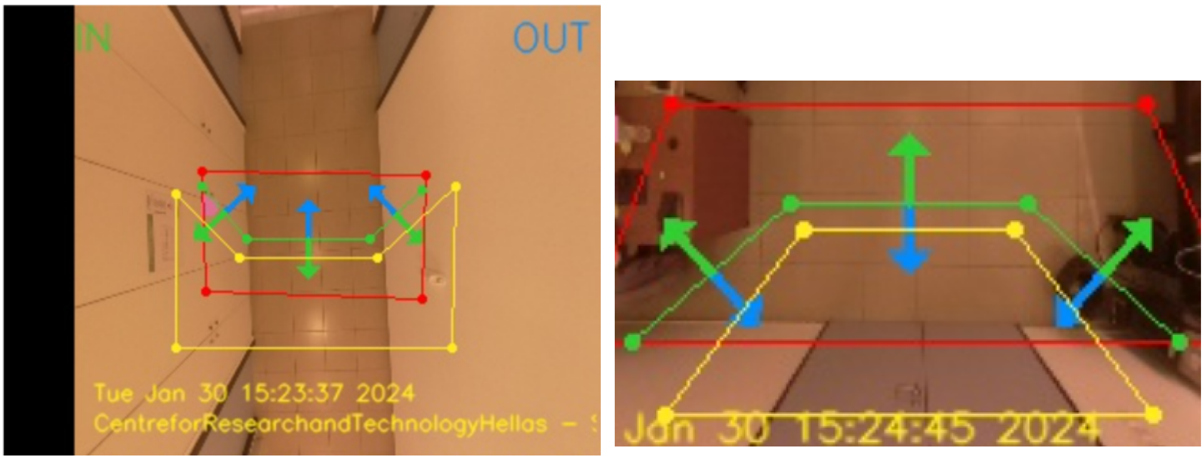


Figure 60. FootfallCam hallway (left) and camera room view (right)

In terms of networking the company offers a straightforward API that can obtain live feed from the camera. Using the internal IP of the sensor and specific parameters we can get the persons coming in and out as dictated in the installation. Initially we need to obtain the access token using a URL as it can be seen in the example below:

```
http://[insert Internal IP]/cgi-bin/access_token.cgi?username=[insert Username] &password=[insert Password]
```

In this way we can get the access token which can be utilised to obtain the data from the device. A command that is employed to get the device data is given below:

```
http://[insert Internal IP]/cgi-bin/apiCount_cgi?data_type=[insert Data Type]&data_format=[insert Data Format]&resolution_min=[insert Resolution Minutes]&date_start=[insert Date Start]&date_end=[insert Date End]&time_start=[insert Time Start]&time_end= [insert Time End]&access_token=[insert Token]
```

We validated the cameras' right operation by having persons going inbound and outbound and we found that the results were less error-prone to calibration. Again we tested the data obtained by the API against the dashboard values with persons crossing the ROI.

The idea is to have the two solutions ready for deployment in order to maximise the number of detection sensors available that will be used for different purposes. These sets will be used for the occupancy while the wireless camera prototype can be utilised for the queuing of the persons in the airport common places.

Indoor air quality sensor operation

The indoor air quality sensor comprised a number of components that provided an integrated system as it can be seen in Figure 61.

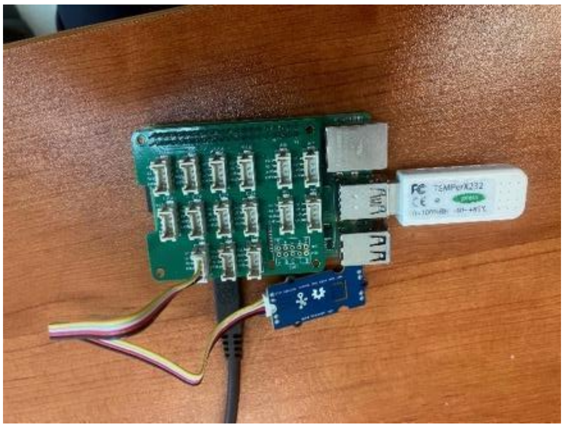


Figure 61. Indoor air quality sensorboard

The sensorboard consists of a raspberry Pi 4 model B as a microcontroller, a TemperX232 USB thermometer and humidity sensor, the Grove-VOC and eCO2 sensor (SGP30) with its respective hat as they are explained in deliverable for the D 4.4.

The validation of the SGP comes as an immediate result of the available code and operability provided by the website of the sensorboard (https://wiki.seeedstudio.com/Grove-VOC_and_eCO2_Gas_Sensor-SGP30/). It has been tested in terms of operations and the results showed satisfactory results, since the finding did not surpass a normal level for an office environment. The code is very simple and a simple unit testing has been introduced whereby print methods have been placed in different parts of the code to see their operations.

For the thermometer we found code on the Internet that provided the simple read of the sensors since we wanted to integrate the readings with the SGP30 in one file. The code can be found in https://github.com/urwen/temper. For validation purposes we utilised the tool that comes with the sensor to measure the temperature and humidity and cross-reference that with the raspberry Pi reading. We found that the two methods coincide and the same readings have been found. Note that we performed testing of the code with print methods to check its efficiency and successful operation. The output of the tool that the TemperX232 provides and we used for validation is given in Figure 62.

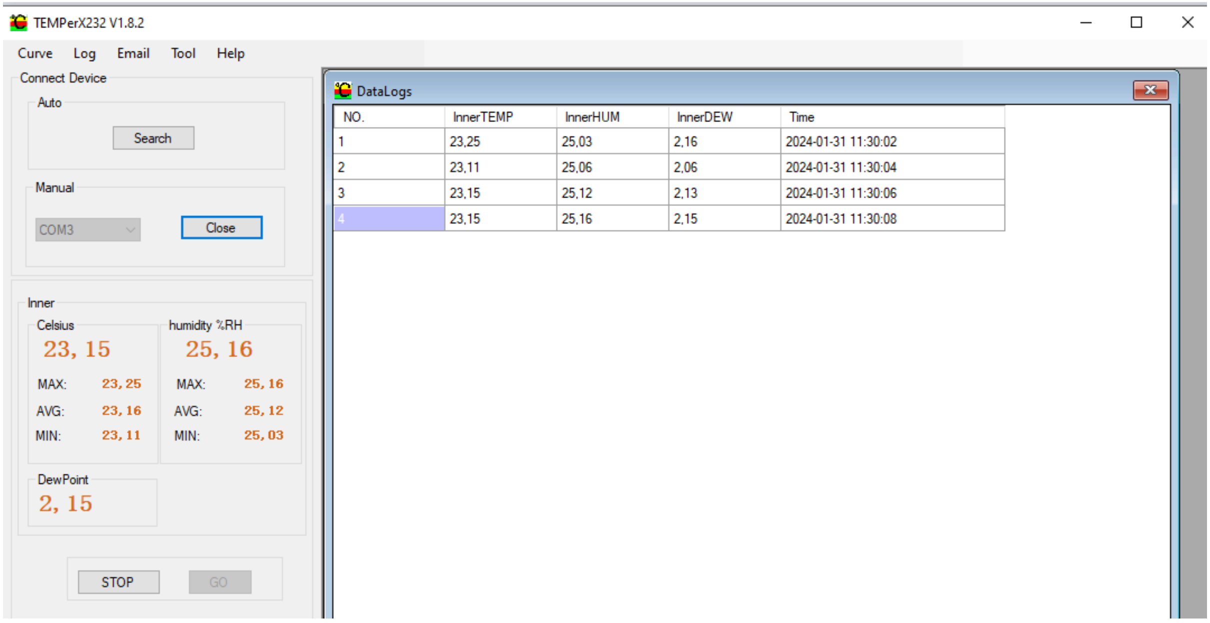


Figure 62. Thermometer and humidity sensor tool output

Software Components Validation

ML Infection Classifier Validation

In the development of our infection ML classifier, we undertook the creation of a comprehensive artificial dataset with the primary objective of discerning and classifying the likelihood of COVID-19 as either high or low. This initiative involved the meticulous design of 27 distinct features, each corresponding to the 9 common places prevalent within our application, namely "Shop1," "Shop2," "Shop3," "Toilet 1," "Toilet 2," "Toilet 3," "Coffee Shop," "Restaurant," and "Cantine." Notably, our application's flexibility allows for the seamless incorporation of additional shops, ensuring adaptability to evolving needs.

To capture the nuanced dynamics of each common place, we structured three features for every location: “<common_place>_coming,” “<common_place>_queue,” and “<common_place>_in.” This detailed approach aimed to provide a comprehensive representation of passenger interactions within these spaces. In addition to defining the features, we strategically introduced a pattern that signified a heightened likelihood of COVID-19, carefully tailored to specific feature values. To account for real-world variability, a percentage of error was incorporated into the dataset. It has to be highlighted that in order to validate our ML classifier we simulated the passengers’ preferences insertion, the person counting and the person queue acquisition using injected data from a script that we developed. In this manner we wanted to simulate the human factors so that we validate that the model was the correct one.

We trained the Random Forest model using 70% of the 2000 entries and we kept the remaining entries for testing. We converted the trained model to a pickle object and we wrote a Python script that gets the last value from a csv file to classify the infection likelihood. This simulates the last record from the database meaning that there is traffic in our airport. An example can be seen Figure 63 for the line where the likelihood is low (0)

```
3,4,1,3,4,6,1,2,7,3,5,3,7,2,5,3,0,2,0,3,0,4,3,5,7,9,7,0
```

```
C:\Users\User\Documents\randomforest>python3 pickled.py
[0]
```

Figure 63. Pickle object result of single csv line

We also performed a different experiment allowing two lines of csv to be present and the result was a list with the correct infection likelihood. These lines are given below and the result in Figure 64:

```
1. 3,4,1,3,4,6,1,2,7,3,5,3,7,2,5,3,0,2,0,3,0,4,3,5,7,9,7,0
2. 15,13,16,19,15,12,11,19,18,16,11,15,17,18,17,18,13,19,14,14,20,18,14,17,16,17,17,1
```

```
C:\Users\User\Documents\randomforest>python3 pickled.py
[0 1]
```

Figure 64. Pickle object result of csv lines

In terms of unit testing the machine learning model creation did not require a thorough examination since it is a very well-known classifier and the pickled script consisted of a few lines of code.

Air Quality Classifier

Since we did not have a dataset of indoor air quality due to the development time of our sensorboard we had to find a way to start working on the tool we envisioned. Our data exploration led us to discover a pertinent dataset available in CSV format on https://github.com/twairball/gams-datasetm. This dataset encompasses 135,100 entries across seven key features: datetime, CO2 levels, humidity, PM10 and PM25 particulate matter concentrations, temperature, and VOC (volatile organic compounds).

Our primary objective was to categorise air quality based on these measurements. To achieve this, we implemented the k-means algorithm with an extensive range of k values. After running the k-means algorithm with various k values, we analysed the resulting elbow method plot as given in deliverable 4.4. The optimal number of clusters, determined by the elbow method, is identified at the point where the plot exhibits a distinct linear form. To that end we aimed to validate the number of distinct air quality indexes that emerged from the dataset. With this dataset we found four distinct clusters and we named them as Low Air Quality: 0, Low to Medium Air Quality: 1, Medium to High Air Quality: 2 and High Air Quality: 3.

Thereafter, we reconstructed the dataset with the respective variable in each line and we ran the random forest algorithm to detect the air quality index based on the object and random numbers as features to investigate. We can see the random forest classifier training accuracy in Figure 65.

```
Model accuracy score with 10 decision-trees : 1.0000
Model accuracy score with 100 decision-trees : 1.0000
```

Figure 65. Random forest accuracy

An example of testing is the following line where the last value is the number 1 or "Low to Medium Air Quality":

```
1. (72,1177.0,60.5,16.2,15.1,25.98,63,1)
```

where we obtained the result "Low to Medium Air Quality". This is a manner that we used to simulate the indoor air quality to determine the index that we will use for our tool.

Queuing and Distance Measurement

The queuing measurement for the weighting factor that has been introduced in WP4 has been easily obtained using the Yolo model, which contains the class of person as object detection. This has been validated in both video and live camera feeds, and the number of persons have been successfully identified. Yolo is a very well-known object detection model, and validated it using persons in our lab. The implementation includes the number of detections that can output the queuing of the people within the range of the camera.

In terms of distance measurement, we obtained the pcap file from the Velodyne VLP 16 lidar and by using the tool that accompanies it, we formed the objects that have been scanned by the sensor. Thereafter we were able to draw the distance that has been detected between two people which coincided with the measurement of the sensor. The reader can find a picture of the measurement in the figure shown in deliverable 4.4 and the respective section.

Weighting Factor Validation

The crux of the routing process lies in validating the weighting factor. Initially, we utilised Alylogic software to simulate the 3rd floor of the HIT as an airport retail area, leveraging the floor plan of HIT offices where offices are equated to stores. As illustrated in the figure within deliverable 4.4, there are three types of stores (A, B, and C), each with a pair of stores. These establishments encompass a variety of amenities typically found in airports, including eateries, cafes, duty-free shops, and restrooms.

Upon entering the store area through any of the entrances, passengers are afforded the opportunity to patronise each of the three store types according to their preferences. It's worth noting that passengers may only visit one store of each type and only once. Upon completing their desired shopping, passengers exit the simulation model.

Every store has a maximum occupancy limit determined by its square footage, calculated by dividing the store's square metres by 4 (equating to 4 square metres per passenger). This aligns with COVID-19 regulations mandating adherence to the 4 square metre rule in indoor spaces.

To execute the simulation, users must furnish input data specifying the number of passengers entering the model, the time interval between each passenger's arrival, and the duration of time they intend to spend in each type of store. These particulars are detailed in Table 27. HIT 3rd floor simulation model input data.

Table 27. HIT 3rd floor simulation model input data

Input data	Value
Number of passengers	1000
Interarrival time for each entrance	25 passenger per hour (50 passenger per hour at total)
Spending time at each store 5 minutes	

It should be noted that the times spent by passengers in each of the stores is not based on real data. However, in the context of the comparison presented in this research, the result is not affected by these values because they are fixed at every simulation.

For the simulation, a list of passengers must be given as input from the user to the model, which will include the preferences (Store A, Store B, Store C) of each passenger and whether they are carriers of COVID-19. In the context of this study, a comparison of the transmission of COVID-19 amongst passengers is made, in the case of passengers visiting the stores in a random order versus visiting the stores in a preset order calculated by a routing algorithm. For that reason, a Python script was developed in order to generate a list of passengers with random preferences and a random order of store visits.

For the simulation, a sample of passengers was given as input to the model from which:

- 70% of passengers will want to visit Store A
- 70% of passengers will want to visit Store B
- 70% of passengers will want to visit Store C

Regarding the percentage of passengers who have COVID-19, simulations were performed with percentages of passengers with infection rate of 2%, 4%, 7%, 10%, 15% and 20%. Also, for each percentage of infected passengers, 40 simulations (20 simulations without passenger routing and 20 simulations with passenger routing) were performed and the results presented below are the averages of the results.

The results of the simulations are given below.. In particular, Figure 66 shows the total exposure time of healthy passengers to COVID-19 inside the store area for the case of the passengers randomly visiting the stores and for the case of the passengers routed to the shops for the different passengers' infection rates. According to the results, it is observed that by using the algorithm to guide the passengers in every case, there is a reduction in the transmission of the virus.

Figure 66. Passenger exposure time at the stores area

The maximum number of passengers per store and simulation with and without routing is shown in Table 28. Maximum number of passengers per store with (blue cells) and without (white cells) routing. Note that the grey columns are with routing and the white without routing. According to Table 28. Maximum number of passengers per store with (blue cells) and without (white cells) routing with the routing of passengers in the airport area with the stores, a decrease in the maximum number of passengers in each store can be seen. For instance, at the Store A (Store A (1) and Store A (2)), for an infected passenger rate of 2%, the average maximum number of passengers in the case without passenger guidance is 6.15 and 5.9 while in the case with passenger routing it is 6 and 4.1 respectively. The results show that passenger navigation inside the store area considerably contributes to preventing congestion in the stores and, as a result, lessens the spread of COVID-19.

Table 28. Maximum number of passengers per store with (blue cells) and without (white cells) routing

Infection rate %	2	4	7	10	15	20
Store A (1)	6.15	6.8	5.6	6.2	6.3	6.65
Store A (1)	6	5.75	5.75	5.85	5.65	5.95
Store A (2)	5.9	5.7	5.95	5.8	5.95	6
Store A (2)	4.1	4.35	4	4.05	4.3	4.5
Store B (1)	6.8	6.1	6.45	7.6	7.15	6.35
Store B (1)	6.35	6.2	5.9	6.15	6.25	6.4
Store B (2)	5	5	5	5	5	5
Store B (2)	3.95	4	4	4.3	4.15	4.2
Store C (1)	6.75	5.85	6.25	6.4	6.4	6.7
Store C (1)	4.4	4.6	4.95	4.5	4.45	4.6
Store C (2)	6	7.2	6.75	6	6.65	6

Store C (2)	4.85	4.85	5.25	5.15	5.05	5.25

Web Service Validation

The idea behind the web service implementation was the interfacing of the Android application with the cloud infrastructure. Hence, since we implemented nine shops, we added a variable for each that would change from 0 to 1 if it was not selected and selected respectively. The application essentially sent a string of 1s and 0s according to which shop was selected. We performed this check using the https://reqbin.com/ website. We added print outs to every connection query execution and we found that there was no error.

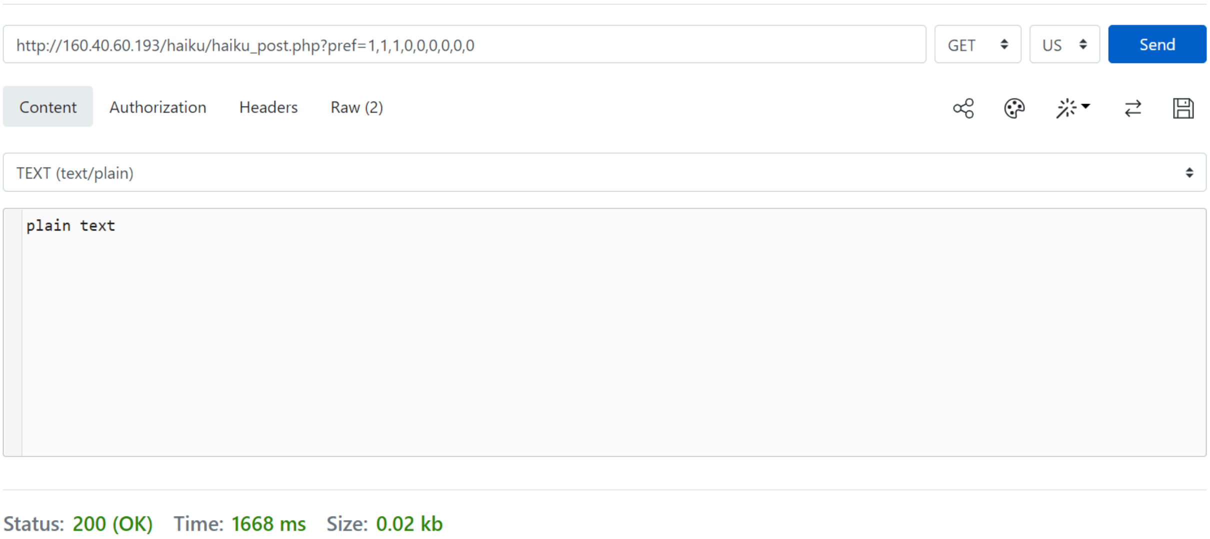


Figure 67. Web interface to send HTTP request to web services

We used the String 1,1,1,0,0,0,0,0,0 to simulate the fact that the passenger wants to visit Shop 1, Shop 2 and Shop 3. We did so with the HTTP GET request that can be seen in Figure . The execution of this request provided the result of Figure 68. There we set a print to show the values of each Shop followed by each Shop. As we can see the values are given in ascending order and the Shops are sorted as well with the priorities given with the order of appearance.

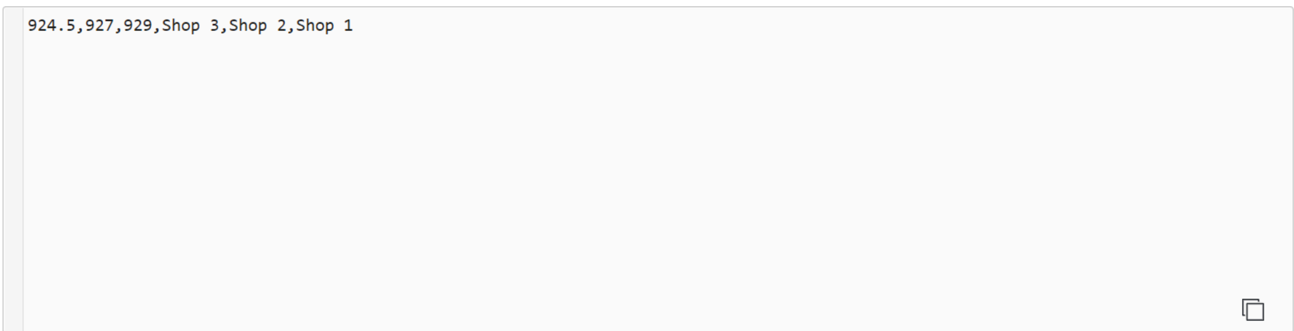


Figure 68. Web interface validation response

That output was for validation purposes. The actual output of the web service can be seen in Figure 69, which is the likelihood of infection (here is 0 meaning low), followed by the sequence of Shops prioritised with the order of appearance. The likelihood takes place with the ML classifier that we explained earlier in the report.

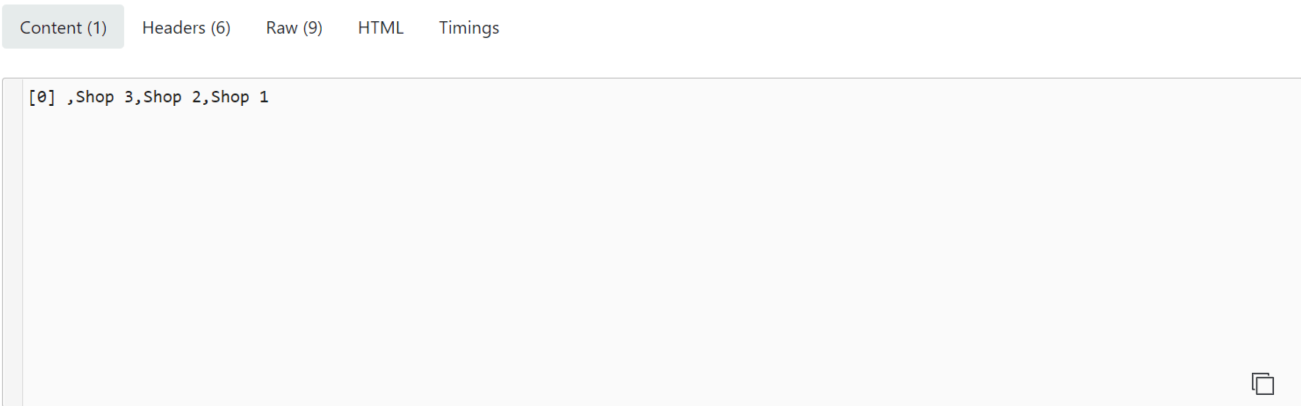


Figure 69. Web interface actual response

Thereafter, we changed the staying time for Shop 2 (all of them were initially set to 30 minutes) We changed it to 60 minutes and we obtained the result in Figure 70. As we can see the order has changed since the algorithm calculated Shop 2 to be the less efficient.

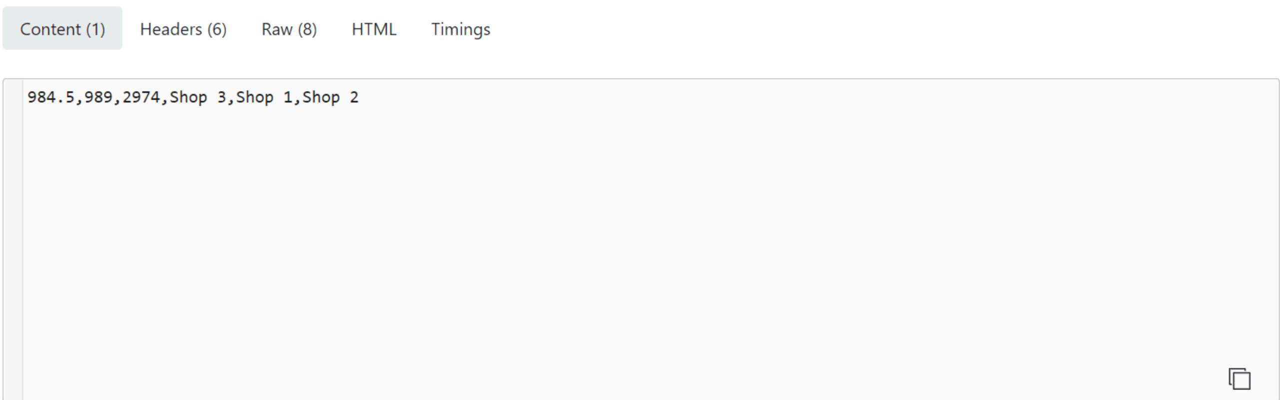


Figure 70. Web interface modified validation response

System Validation

This section comprises the overall system validation of the two developed tools namely the Android application and the statistical and explainability tool for air quality.

Android Application Validation

The third validation test was based on the web service validation using the online tool and was the Android application validation. A brief operation of the Android app is given in the video here. Initially we validated the fact that the preferences coming from the preference screen were successfully transferred to the notification screen where the output was displayed. We used the same shop preference with the online validation choosing Shop 1, Shop 2 and Shop 3 as it can be seen in Figure 71.

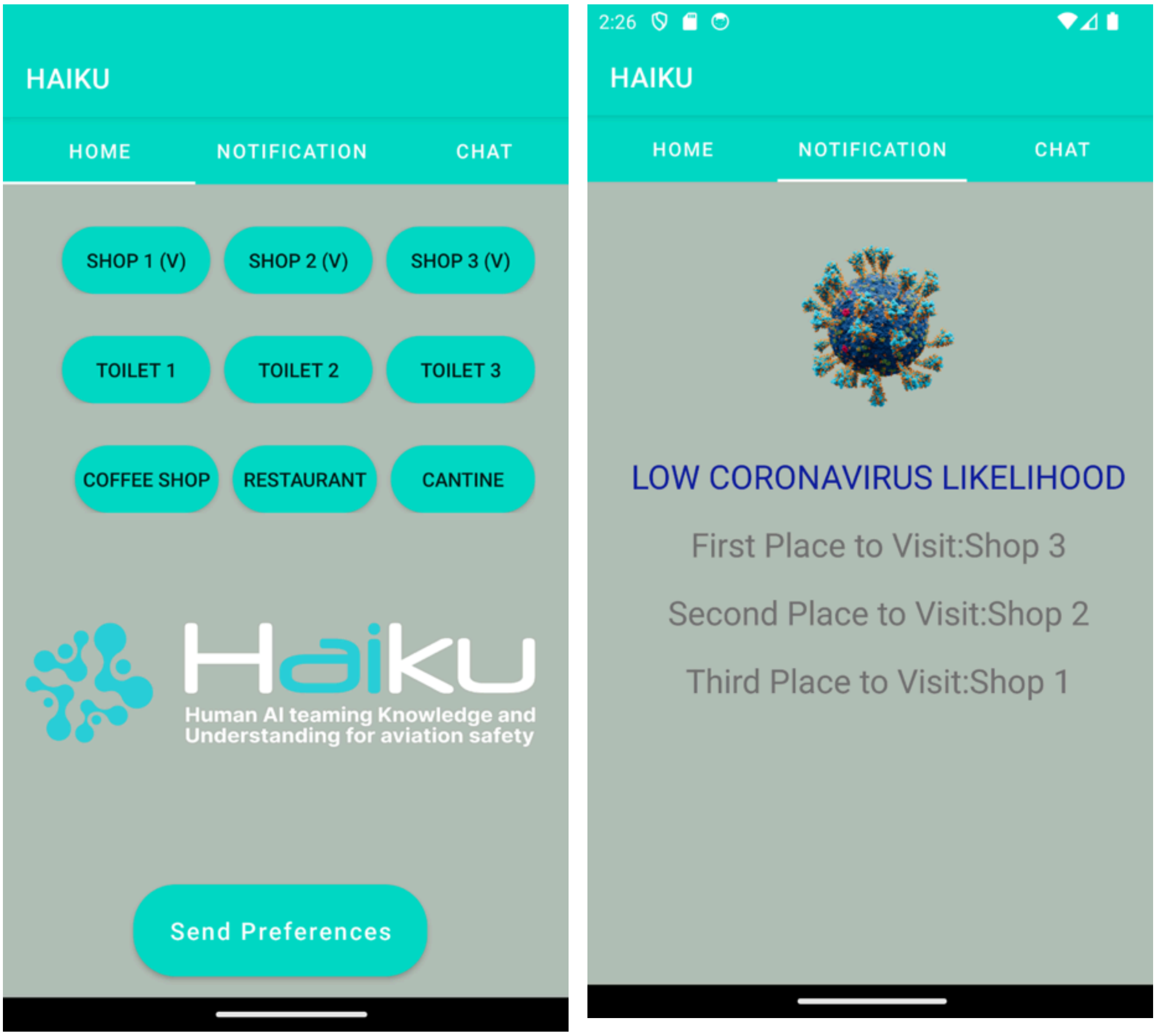


Figure 71. Validation preferences of application and output of notification tab

We performed unit testing on the application using Logs that appear in the Logcat interface of Android Studio. In this way we made sure that the correct string was transferred to the correct URL and that the response data coincided with the ones shown in the online web service validation shown before. The output of the Logcat can be seen in Figure 72.

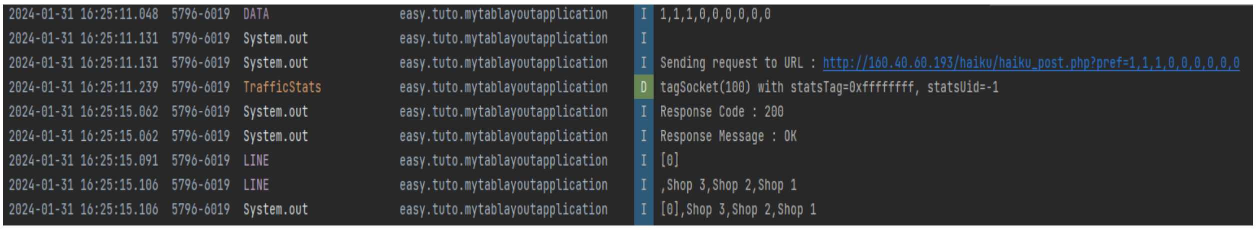


Figure 72. Logcat output

As we can see the correct string is transferred after the notification fragment is displayed and the URL GET request is shown as well as the response coming from the web service hosting the weighting factor algorithm. In this way we can validate the response shown in the notification tab which is shown in Figure 71 and coincides with the request made previously.

Chat-like Bot and Passenger Communication

The passenger communication takes place with the chat tab where the passenger can interact with the bot to get valuable information regarding the routing inside the airport. The way this was implemented was by using a SQLite Database to store the questions and potential answers given by the bot. Initially the bot introduces itself on the passenger as it can be seen in Figure 73.

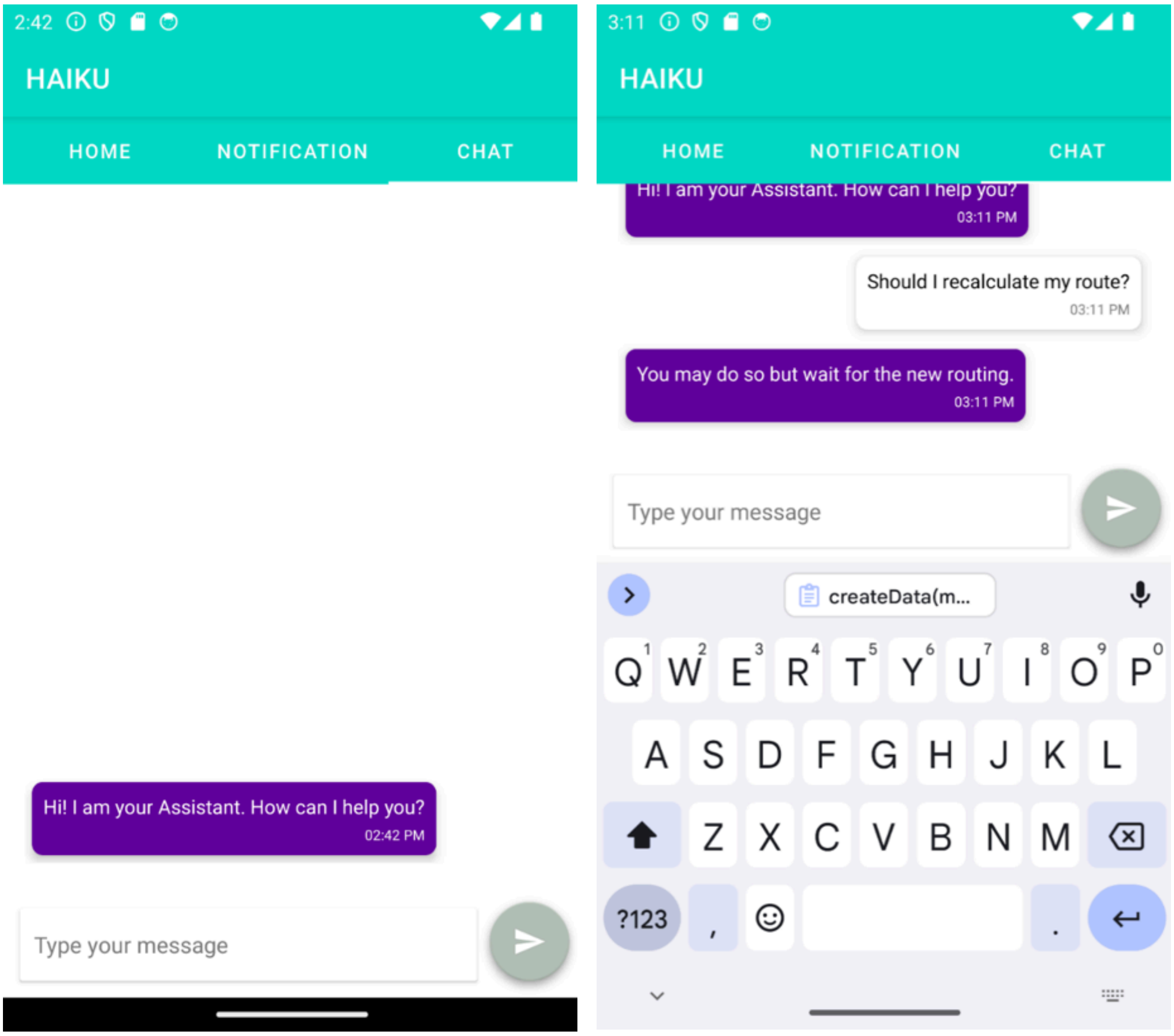


Figure 73. Chat introduction message and Q&A between passengers and the bot

We were able to validate the Knowledge Base by inserting more Q&As and getting the responses by the bot. The reader can see a Q&A between the passenger and the bot in Figure 73. Notably the Knowledge base can learn from the passenger to react accordingly to a similar question.

The reader can see the reaction to unknown questions raised to the bot in Figure 74. There the bot asks how to respond to this question. In this manner we aim to populate the Knowledge Base with more Q&As in the same manner that state-of-the-art AIs perform.

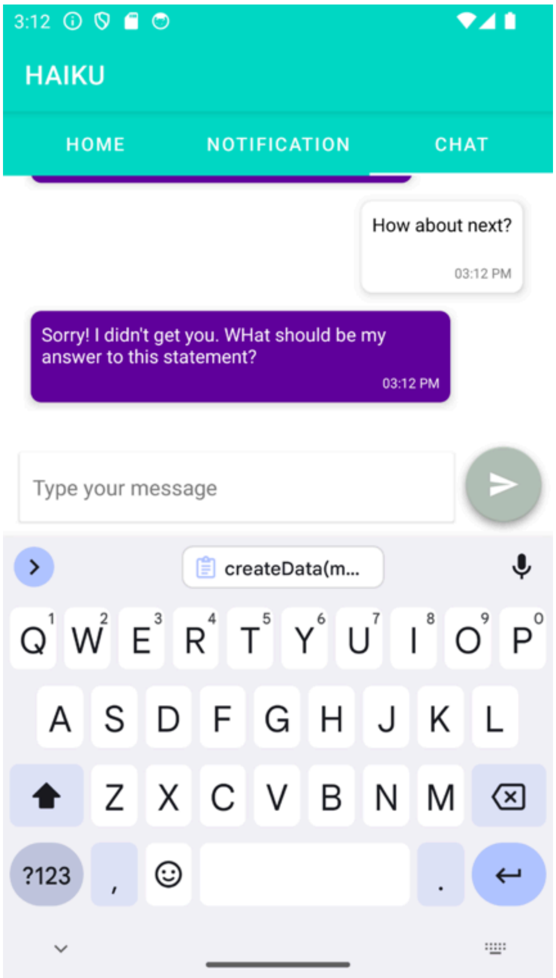


Figure 74. Bot reacting to an unknown question

We managed to validate some of the functionalities of the communication between the passenger and the Android bot. We were able to investigate the insertions on the Knowledge Base and the respective responses.

The idea was to populate a Knowledge Base based on the responses of the weighting factor and implement the CLT levels as notifications coming to the passenger by the AI. The first CLT has been implemented in the notifications tab by showing the first explanation of the likelihood index of the COVID-19. The chat as it currently stands represents the second stage of the CLT with some explanation given to the passenger. CLT levels have not been implemented yet and they are going to take place using an internal timer on the phone that will take the response from the weighting factor and talk to the passenger according to the times given in the CLTs. The implementation and validation will take place in the immediate future.

Statistical and Explainability Tool

As seen in deliverable for D4.4 we developed a statistical tool that also presents the air quality index and an explainability set of plots using the SHAP method. A demonstration of the tool can be seen at https://shorturl.at/eirzK. The tool has been implemented using the Stremlit framework, which is an easy Python module framework that creates web services and portals, with the purpose of placing it to our cloud infrastructure.

We validated that the selected value from the dropdown box corresponds to the values coming from the csv using the portal by inspecting the DataFrames that are built to populate the plot, which is also shown in the respective section of the deliverable 4.4.

The next part of the tool is the air quality index and the explainability part of the UI. The air quality index has been validated using the classifier as a standalone Python module as explained earlier in the report and the output has been validated using the ML pickled model. The output is a simple label that is provided by the Streamlit framework. The explainability algorithms have been initially validated using the SHAP method, which is a very well-known explainability method. We ran an experiment that outputs the Shapley values for the particular dataset and we found that they coincide with the result from the tool. We initially run the Python code as standalone, with the first plot being validated against the value of the csv line from the pickled model. The second plot used the entire training set so the code has been tested before integrating it to the portal.

8.4 UC#6 Validation 1 Conclusion

Upon completion of VAL1, these are the main lessons learned that point to new design directions:

Table 29. UC#6 lessons learned

Insight	Functional Requirements	Proposed solution for next iteration
The validation of the components were essential to build the overall systems that were presented in this UC.	The components were tested in real time and they exhibited good overall results	Use more devices and other ML algorithms for comparison.
The devices were developed to be used as a developed network.	There have to be networked applications that will connect the lot of the devices to the same system	Building more devices
The Android application is the system that comprises a number of tools to output the routing and the likelihood of infection	The tool works in real time and maximum number of connections need to be tested	Refine the app and make it more user-friendly rather than just fully functional
Statistical and explainability tool for air quality was essential for the health and covid relation	This needs to provide a correlation between the infection likelihood and the air quality which needs to be done in training and evaluated in real time	We need to get data from cameras and mobile phones to proceed with correlation.

Furthermore, we discovered that we need to validate the training sets with a level of uncertainty, due to the fact that there could be noise on the data. In terms of the AI model, we will consider the use of other algorithms that may be used using the TPOT model which checks the most appropriate ML model for our data using genetic algorithms. We also believe that the CLT concept should be transformed to an algorithm.

In terms of the Android application, we think that refining the HMI could bring new insights in the interaction between the passenger and the system. A better and more engaging knowledge base is the key for the success of our application. The statistical tool needs to be refined also and include more data on-the-fly according to the current implementation.

The validation of the components as well as the system gave us the insight that we will have minor problems in the deployment of the system in the Amygdaleona airport in Kavala. The systematic validation provided no significant issues; hence, we will concentrate on the refining of the pilot and the further investigation of the pilot case we will consider in the airport. Finally, we are eager to validate the system as a whole with real human subjects as passengers, which will provide the desired TRL5/6 to our UC.

9. General conclusion

The objectives of Validation 1 activities varied among Use Cases and depended on their respective levels of advancement. For all UCs, the primary objectives were to:

1. Collect feedback from end-users,
2. Collect evidence of the operational relevance and feasibility of the AI.

Collectively, the results from all UCs help us derive a better understanding of the challenges and solutions in achieving human-AI teaming in aviation in general, as seen through the lens of our Human-AI Teaming (HAT) framework presented in D4.1 Operational system / technical requirements.

Table 30 overviews the **key differences in validation approaches** across the six UCs. From this table it can be seen that UCs address both real-time operational phases and planning and post operations. The goal of the **teaming effort varies between UC**, covering a range of aspects including restoring human situation awareness, improving decision making efficiency, and enhancing safety. Furthermore, the type of intelligent assistants conceptualised and explored in VAL1, along with its capabilities and authority, varies across UCs. Regarding **EASA AI Level classification**, UCs have explored different levels within the spectrum of human assistance (1B), cooperative (2A), and collaborative (2B). Finally, variations between UCs are apparent with respect to the **HAT requirements considered in VAL1**. All UCs have explored operational performance and teamwork/collaboration. Most have also explored explainability/transparency and situation awareness. Only a few have explored workload and safety. None have so far explored ethics or information security.

Table 30: Use Case differences in VAL1 validation approach.

Use Case	Aviation segment and operational phase	Goal for human-AI team	End user	Type of Intelligent Assistant (EASA classification)	HAT requirements considered
#1 Flightdeck startle response	Airborne (in flight)	For the human to recover from the startle & surprise effect.	Pilot	Adviser, Monitoring partner (1B, 2A)	Operational performance; Teamwork/collaboration
#2 Flightdeck route planning	Airborne (in flight)	Determine a suitable alternate destination for rerouting due to deteriorating weather conditions.	Pilot	Team member, Decision Making Support (1B, 2A, 2B)	Operational performance; Teamwork/collaboration; Situation awareness; Explainability/transparency
#3 Urban Air Mobility	U-space traffic managemen	Managing aircraft rerouting within U-space during normal and emergency	UAM Coordinator	Team member, Monitoring nominal & off-nominal	Operational performance; Teamwork/collaboration; Situation

	t (during operations)	situations, which involves coordinating with external stakeholders.		situations, Adviser (2B)	awareness; Explainability/transparency; Safety
#4 Digital Tower	Airport ATM (during operations)	Determine an optimal runway utilisation by determining the best sequences of arriving and departing aircraft.	Tower Controller	Team member, Pre-processing information for humans (1B)	Operational performance; Teamwork/collaboration; Situation awareness; Explainability/transparency; Safety; Workload
#5 Airport safety management	Airport (planning and post-operations)	Reducing the risks of three key incident types: pushback error, hold-point busts and incorrect taxiway selection.	Airport safety duty holder & Stack Partners	Assistant for data analytics (1B)	Operational performance; Teamwork/collaboration; Situation awareness; Explainability/transparency
#6 Airport COVID-19 monitoring	Airport (during operations)	Mitigating the risk of spreading contagious diseases within an airport.	Passengers & airport staff	Assistant for airport staff and visitors (1B)	Operational performance; Teamwork/collaboration; Explainability/transparency

Each UC was able to identify ways to improve its assistant.

- Regarding **explainability**, it was important for UC 1, 2, and 4 to present concise and accurate information due to the complex environments in which the users (Pilots and ATCos) work. End users were often overwhelmed with unnecessary information, so the adoption of a common explainability framework was instrumental in steering the right design choices. In contrast, UC3 users wanted more explanations about the outputs of the assistant.
- UC1 and UC4 involved validation in actual operational scenarios. For these UCs, the assistants were most useful **when the complexity of the scenarios increased and less useful in normal situations**.

The second main objective of UCs validation was to evaluate the integration and performance of the assistant components.

- UC1: the goal was to integrate the assistant into a highly standardised environment, namely the flight deck.

- UC5: the data from Luton Airport were evaluated in terms of volume and quality.

- UC6: the focus was on validating each component of the assistant. UC6 assistant components were tested in real-time and in realistic environments.

- UC3: The focus was on evaluating end users' perceptions, understanding, and interaction experiences with the intelligent assistant to gather feedback on its concept of operations. The implementation of the intelligent assistant for VAL1 emphasised the human-machine interface, while the backend system was operated using a wizard of Oz method.

In conclusion, Validation 1 was a necessary step for all UCs at the project's midpoint to ensure a human-centric development approach by considering end-users feedback while testing the prototypes. The objectives of validating the operational relevance and key technical component performances of the assistants were essential for evaluating Human-Machine teaming concepts in the next validation phase (Validation 2).

10. References

Boeing. (2022). Concept of Operations for Uncrewed Urban Air Mobility.

Choi, K. Y., & Ishii, H. (2020). ambienBeat: Wrist-worn Mobile Tactile Biofeedback for Heart Rate Rhythmic Regulation. Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction, 17–30. https://doi.org/10.1145/3374920.3374938

Costa, J., Adams, A. T., Jung, M. F., Guimbretière, F., & Choudhury, T. (2016). EmotionCheck: Leveraging bodily signals and false feedback to regulate our emotions. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 758–769. https://doi.org/10.1145/2971648.2971752

EASA. (2023). Guidance for Level 1 & 2 machine learning applications. Technical Report, EASA. https://www.easa.europa.eu/en/downloads/137631/en

Elgin, P. D., & Thomas, R. P. (2004). An Integrated Decision-Making Model for Categorizing Weather Products and Decision Aids. https://ntrs.nasa.gov/citations/20040040165

EVE Air Mobility. (2022). Concept of Operations for sustainable urban air mobility in Rio De Janeiro. Brazil.

FAA. (2020). Unmanned Aircraft System (UAS) Traffic Management (UTM) Concept of Operations v2.0 Washington, USA: Federal Aviation Administration.

Federal Office of Civil Aviation (FOCA). (2020). Swiss U-Space ConOps.

Hourlier, Sylvain, Jaime Diaz-Pineda, Marc Ganti, Aurélien Thiriet, and Daniel Hauret. (2022). "Enhanced dialog for Human-Autonomy Teaming - A breakthrough approach." IEEE/AIAA 41st Digital Avionics Systems Conference (DASC). IEEE. doi:10.1109/DASC55683.2022.9925736.

Lundberg, J., Arvola, M., & Palmerius, K. L. (2021). Human Autonomy in Future Drone Traffic: Joint Human–AI Control in Temporal Cognitive Work. Frontiers in Artificial Intelligence, 4. https://www.frontiersin.org/articles/10.3389/frai.2021.704082

McCraty, R., & Zayas, M. A. (2014). Cardiac coherence, self-regulation, autonomic stability, and psychosocial well-being. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.01090

Perciavalle, V., Blandini, M., Fecarotta, P., Buscemi, A., Di Corrado, D., Bertolo, L., Fichera, F., & Coco, M. (2017). The role of deep breathing on stress. Neurological Sciences, 38(3), 451–458. https://doi.org/10.1007/s10072-016-2790-8

Schlatter, S. T., Thérond, C. C., Guillot, A., Louisy, S. P., Duclos, A., Lehot, J.-J., Rimmelé, T., Debarnot, U. S., & Lilot, M. E. (2022). Effects of relaxing breathing paired with cardiac biofeedback on performance and relaxation during critical simulated situations: A prospective randomized controlled trial. BMC Medical Education, 22(1), 422. https://doi.org/10.1186/s12909-022-03420-9

SESAR JU. (2022). CORUS-XUAM. https://corus-xuam.eu/

Sun, Y., Lu, T., Wang, X., Chen, W., Chen, S., Chen, H., & Zheng, J. (2023). Physiological feedback technology for real-time emotion regulation: A systematic review. Frontiers in Psychology, 14, 1182667. https://doi.org/10.3389/fpsyg.2023.1182667

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. Psychological Review, 117(2), 440–463. https://doi.org/10.1037/a0018963

Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research Part C: Emerging Technologies, 5(1), 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

Yu, B., Funk, M., Hu, J., Wang, Q., & Feijs, L. (2018). Biofeedback for Everyday Stress Management: A Systematic Review. Frontiers in ICT, 5, 23. https://doi.org/10.3389/fict.2018.00023