Haiku
Human AI teaming Knowledge and Understanding for aviation safety

Deliverable N. 7.3

Validation of the SHS case-based approach in case studies

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

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

Abstract:

The human-centric design approach is a key focus of the HAIKU project, which aims to integrate societal, ethical, and value-based insights into AI design. This project specifically centres on advancing Intelligent Assistants (IAs) as a type of AI-powered solution. The validation process revolves around diverse aviation scenarios, encompassing airport management, Air Traffic Management (ATM), and flight operations.

To pursue this objective of developing a human-centric design approach, this deliverable presents the outcome of a proactive assessment of Safety, Human Factors (HF), Security, and Liability (SHS-L) issues associated with the development and deployment of AI-based Intelligent Assistants. The primary objective is supporting and refining the concepts proposed by the HAIKU use cases (UCs).

The validation approach takes a systematic perspective, encompassing five essential components: safety, security, human performance, liability, and regulation. The lessons learned from this assessment serve as valuable input to enhance the HAIKU validation framework, resulting in recommendations and conclusions that drive updates for individual UCs and the overall project development.

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

Information table

Deliverable Number	7.3
Deliverable Title	Validation of the SHS case-based approach in case studies
Version	1.1
Status	Final
Responsible Partner	DBL
Contributors	Paola Lanzi, Nikolas Giampaolo, Elisa Spiller
Contractual Date of Delivery	31.08.2023 (M12)
Actual Date of Delivery	31.08.2023 (M12)
Dissemination Level	PP

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

Document history

Version	Date	Status	Author	Description
0.1	22.06.2023	Draft	Paola Lanzi (DBL) Nikolas Giampaolo (DBL) Elisa Spiller (DBL)	ToC
0.2	11.07.2023	Draft	Paola Lanzi (DBL) Nikolas Giampaolo (DBL) Elisa Spiller (DBL)	Draft
0.3	18.07.2023	Draft	Paola Lanzi (DBL) Nikolas Giampaolo (DBL) Elisa Spiller (DBL)	Internal version
1.0	31.07.2023	Draft for internal review	Paola Lanzi (DBL) Nikolas Giampaolo (DBL) Elisa Spiller (DBL)	Consortium revision
1.1	31.08.2023	Final version consolidated further to internal review by ECTL, DBL	Paola Lanzi (DBL) Nikolas Giampaolo (DBL) Elisa Spiller (DBL)	Final version

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

List of acronyms

Acronym	Definition
AI	Artificial Intelligence
AMC	Acceptable Mean(s) of Compliance
ANSP(s)	Air National Service Provider(s)
ASW	Airport Safety Watch
ATC	Air Traffic Control
ATCO	Air Traffic Control Officer
ATM	Air Traffic Management
BFU	German Federal Bureau of Aircraft Accidents Investigation
CART	Council Aviation Recovery Task Force
CAT	Commercial Air Transport
CIS	Common Information Services
COMBI	Bidirectional Communicator
CONOPS	CONcept of OPerationS
(D)DoS	(Distributed) Denial-of-Service
DUC	Digital assistant for UAM Coordinator
e.g.,	exempli gratia (for example)
i.e.	id est (that is)
EASA	European Aviation Safety Agency
EC	European Commission
EU	European Union
HAIKU	Human AI teaming Knowledge and Understanding for aviation safety

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

HAT	Human AI Teaming
HF	Human Factors
HMI	Human-Machine Interface
IA	Intelligent Assistant
ICAO	International Civil Aviation Organisation
IDPS	Intrusion Detection and Prevention System
ISA	Intelligent Sequence Assistant
KPA(s)	Key Performance Area(s)
LLA	London Luton Airport
LOC-I	Loss-Of-Control In-flight
MFA	Multi-Factor Authentication
MOC	Mean Of Compliance
N/A	Not Available
NATS	National Air Traffic Service
PHC	Public Health Corridors
PIC	Pilot-In-Command
SES	Single European Sky
SHS	Safety, Human Factors and Security
SHS-L	Safety, Human Factors, Security and Liability
UAM	Urban Air Mobility
UAMC	Urban Air Mobility Coordinator
UAS	Unmanned Aircraft System
UC(s)	Use Case(s)

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

US	U-Space
USS	U-Space Services
USSP	U-Space Service Provider
UTM	U-Space Traffic Management
Wi-Fi	Wireless Fidelity

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

Table of contents

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

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

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

List of figures

List of tables

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

1. Introduction

1.1. Intended readership

This document is the first release of *D7.3 – Validation of the SHS case-based approach in case studies*, and reports the results obtained in *T7.3 – Safety, Security and Human Factors analysis*.

The aim of the research done in this task is to **feed and support the design and consolidation of the concepts proposed by the HAIKU use cases (UCs), by performing a preliminary and proactive assessment of possible Safety, Human Factors, Security and Liability (SHS-L) issues** related to the development and deployment of the Artificial Intelligence (AI)-based Intelligent Assistants (IAs) addressed in the UCs. The results of the analysis are followed by a set of recommendations suggesting preliminary adjustments to mitigate the risk exposures for all the actors involved.

This edition – delivered by M12 (August 2023) – is an iterative live document primarily addressed to the HAIKU Consortium. This is the reason why the dissemination level is limited to project partners only. The final version of the report will have a public dissemination level and is expected by M36 (August, 2025).

1.2. Related tasks and documents

The document takes into account some of the deliverables already produced by the HAIKU project, as well as the preliminary results obtained from performing some related tasks.

More specifically, the following submitted deliverables were the main ones taken into account:

- **D3.1 – Human-AI Teaming Framework and Design Document**, which proposes a checklist-driven set of guidelines that can be used to ensure that Human AI Teaming (HAT) principles are adequately addressed in use case design.
- **D3.2 – Concepts of Intelligent Assistants**, which reports the HAIKU concepts, describing the key elements at the strategic level.
- **D3.3 – Human-AI Teaming Validation Framework**, providing the provisional framework for validating use case prototypes, including validation success criteria, constructs, metrics, measurement methods, instruments and protocols.
- **D7.1 - State of the art in Safety, Human Factors, and Security (SHS) assurance processes in aviation**, which presents the legal framework of the HAIKU project and the state of the art in regulations, consensus-based industry standards and best practices, concerning SHS.
- **D7.2 – Development of safety, HF and security approaches for Human Intelligent Assistance Systems**, which provides the assessment methodology and the Acceptable Means of Compliance to be applied and validated in HAIKU UCs.

Considering the tasks running in parallel with T7.3, the document also reflects the preliminary results obtained by the following research activities:

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

- **T2.3 – Analysis of Societal Impact** [M5-M36], which carries out the societal impact analysis for the proposed Intelligent assistants' concepts.
- **T6.1 – Scenario design for each use case** [M7-M36], which aims to ensure the successful engagement of end-users and stakeholders.
- **T6.2 Refine operational concepts of Intelligent assistants' concepts** [M8-M12], which will refine the initial concept of operations for Intelligent assistants in each use case, building on the concepts defined in WP3.

1.3. Structure of the document

The document is structured in 9 sections:

- Section 1 – **Introduction** – is the present introduction [§ 1].
- Section 2 – **HAIKU validation framework** – provides some insights about the methodologies adopted for this first HHS-L assessment [§ 2].
- Sections from 3 to 8 present the **results of the SHS-L assessment** to each UC (respectively, UC1 [§ 3]; UC2 [§ 4]; UC3 [§ 5]; UC4 [§ 6]; UC5 [§ 7]; UC6 [§ 8]).
- Section 9 – **Final considerations** – eventually elicits some final considerations [§ 9].

The report also includes 8 Annexes, respectively containing:

- references and bibliography (Annex A)
- the liability legal framework applicable to HAIKU (Annex B)
- the assessment grids containing the answers provided by the Us owners (correspondingly Annexes C to H)

2. HAIKU design and validation framework

One of the main goals of HAIKU is to develop a human-centric value-based design approach, bringing societal, value-based, ethical insights into the AI design. In this project, the attention converges on advancing a specific type of AI-power solution, i.e., Intelligent Assistants (IAs). In particular, the use cases consider different aviation scenarios, respectively covering airport management, ATM, UAM and flight operations.

The analysis is based on the first version of the project design and validation framework, as described in HAIKU D7.2 issued in August 2023. As illustrated by the figure below (Figure 1), **the purpose of this framework is to develop a systematic human-centric approach to design and validation based on 5 essential key performance areas (KPAs), namely safety, security, human factors (HF), liability and legal compliance**. For each of these KPAs, dedicated assessment methods and tools are proposed that were applied for the analysis reported in this document. The results and the recommendations emerging from the assessment are expected to be used for the design and consolidation of the HAIKU UCs by the respective UCs' owners.

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

Figure 1. HAIKU validation framework

3. UC1 – IA for the flight deck startle response

3.1. Concept description and possible scenarios

The UC1 considers the development of an IA able to enhance the global performance of Commercial Air Transport (CAT) pilots in managing the startle response, in particular in single pilot operations. It addresses the cooperative relationship between pilots and **the AI-based Intelligent assistant**, which **aims to enhance the overall performance of CAT and support pilots in managing their startle response, particularly in unexpected and intense situations**. The purpose of the IA is to support pilots in coping efficiently with startle and surprise reactions, ultimately improving flight safety and preventing loss-of-control in-flight accidents.

As explained in D3.2, the startle effect can be defined as the first response to a sudden, intense stimulus. It triggers an involuntary physiological reflex, such as blinking of the eyes, an increased heart rate and an increased tension of the muscles (Koch, 1999). On the flight deck, the startle effect is often combined with a surprise that results from a disparity between a person's expectations and what is actually perceived (Horstmann, 2006). As the flight deck is the interface between highly automated complex systems and pilots, such disparity between the reality and crew members' expectations can have significant consequences on the safety of the flight. Startle and surprise reactions have played a key role in a significant number of accidents, including Loss-of-Control In-flight (LOC-I).

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

The IA for the startle response collaborates with the pilot to make sense of the situation, coping efficiently with unexpected scenarios in the cockpit while quickly recovering from the deleterious effects of startle and/or surprise. The tool constantly elaborates aircraft data as well as pilot's physiological and behavioural data (e.g., respiration rate, heart rate, gaze position). If unexpected events – like system failure, bird strike, lightning strike or automation surprise – trigger a startle and/or surprise effect for the pilot, the IA is activated. By means of visual and auditory stimuli, the IA improves the pilot's global situation awareness and supports the quick recovery from a startle/surprise.

Based on the information available at the current stage of the design process, the IA would be responsible for four main tasks:

- Startle effect detection
- Short term actions support to help pilots perform the necessary actions to stabilise the situation on unexpected events onset
- Emotion regulation to support pilots' physiological recovery in an efficient manner
- Sensemaking to ensure pilots get all the necessary information to make the appropriate decision.

In light of the above, the AI assistant is asked to perform different roles. The table below (Table 1) reports the Human-AI Teaming Types & Digital Assistants categories developed in HAIKU WP3 (D3.2).

Table 1. HAIKU HAT classification

	To analyse	To manage	To act
	A digital assistant providing information to the user by capturing, processing, and analysing data...	*A digital assistant supporting the user in managing the workflow, organising and prioritising tasks...*	*A digital assistant capable of performing actions/tasks (to face a situation or recover from errors) ...*
...on-demand	**Observer**	**Secretary**	**Rescuer**
...proactively	**Informer**	**Coordinator**	**Executor**

According to this classification frame, the tool should act as:

- informer,
- coordinator,
- executor,
- rescuer.

When acting as **informer**, **coordinator** and **executor** the tool proactively supports the pilot in managing the situation. Instead, when it collaborates with the pilot to make sense of the situation,

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

the assistant could also be considered a **rescuer**.

3.2. HF assessment

During the preliminary HF assessment, we delved into the cooperative relationship between pilots and the AI-based Intelligent assistant.

The Intelligent assistant in this use case is fixed once deployed, and there is moderate cooperation between the developers and the system. The developers collect data to continuously improve the model. However, the specific goals of the human-AI cooperation are still being determined, making it somewhat unclear at this stage. The cooperation aims to support the pilot in managing the startle response, providing short-term actions to stabilise unexpected situations and helping the pilot recover from startle or surprise. There is a focus on enhancing the pilot's situation awareness and providing biofeedback for emotional and physiological recovery.

The interaction between the human and the Intelligent assistant is a one-time engagement triggered by the detection of startle or surprise**. The AI system proactively supports the pilot in real-time, ensuring concurrent interaction. The degree of agency is balanced, with the ultimate decision-making power resting with the pilot.**

The Intelligent assistant is not used by other parties, and the human and Intelligent assistant agents are physically co-located in the cockpit. The pilot is fully aware of interacting with the Intelligent assistant system, and explicit consent is not required before interaction.

Since the Intelligent assistant's purpose is to support rather than replace the pilot's decision-making, the consequences of its failure should be minimal. The potential benefits of the Intelligent assistant depend on the specific situation, ranging from improved situation awareness to critical support during critical flight phases.

Cooperation between the humans and AI could lead to better situation awareness, fewer crashes, and an introduction of AI support in the cockpit. However, users might have concerns about the use of physiological data for assistance. Trust between the human and the Intelligent assistant is essential, and false positives or negatives should be avoided.

The Intelligent assistant interacts with the human through a screen, and its performance should be predictable. At present, there are no confidence indicators for the detection of startle effects, but their implementation is under consideration.

The target group comprises adults devoid of exceptional needs or distinctive attributes, displaying high technological proficiency. Additionally, this group exhibits a diverse array of national and cultural origins.

To ensure the successful and safe cooperation between pilots and the AI-based Intelligent assistant, addressing potential issues and challenges is crucial. **One significant challenge is the need to refine ambiguous goals and objectives through user studies and pilot involvement in the development**

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

process. This lack of clarity may affect pilots' trust in the system and their willingness to rely on its assistance. Therefore, comprehensive training should be provided to pilots to understand the AI system's limitations and potential errors, preventing overreliance on the Intelligent assistant.

Additionally, **there is a concern about pilots becoming overly dependent on the Intelligent assistant, leading to complacency and reduced vigilance**. To mitigate this, fail-safe mechanisms should be implemented to encourage pilots to maintain an active role in critical decision-making processes. The AI system should provide suggestions and assistance rather than taking complete control. Training scenarios simulating situations where the Intelligent assistant might fail or provide inaccurate information can prepare pilots for such occurrences.

To enhance the Intelligent assistant's reliability and predictability, confidence indicators for startle effect detection should be developed. Cross-cultural considerations should also be considered during design, with user training tailored to accommodate different backgrounds and levels of experience.

Some pilots may be apprehensive about accepting AI support in the cockpit, fearing a loss of control or reduced job security. To alleviate these concerns, **pilots should be actively involved in the development and decision-making process regarding the AI system**. Engaging them as key stakeholders and addressing their concerns throughout the design and implementation phases will help build trust and understanding. Emphasising that the Intelligent assistant is intended to enhance their performance and safety, not substitute their role, can further increase acceptance.

By acknowledging and addressing these potential issues, the cooperative relationship between pilots and the AI-based Intelligent assistant can be enriched, improving situational awareness, and more effective management of startle responses in CAT.

3.3. Safety assessment

The safety assessment for the AI-based Intelligent assistant in CAT is currently in the initial design phase. However, several potential safety risks and challenges need to be addressed to ensure the system's safe and effective operation.

Under normal conditions, one potential safety risk is the risk of **overreliance on the Intelligent assistant**. Excessive reliance on the AI system may lead to complacency among pilots, reducing their situational awareness and readiness to take immediate control in case of AI system errors or failures. To mitigate this risk, fail-safe mechanisms and decision support guidelines should be implemented. These mechanisms should encourage pilots to remain actively engaged in critical decision-making processes, ensuring that the intelligent assistant's role is clearly defined as a supportive tool rather than a replacement for the pilot's judgement.

Another potential safety risk under normal conditions is the **unpredictability of the intelligent assistant's performance** and decision-making. If the AI system's behaviour is inconsistent or its decisions are unpredictable, it may lead to unexpected outcomes and undermine pilot confidence in the AI system. Thorough testing and validation of the intelligent assistant system are essential to

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

ensure reliable performance across various scenarios. Additionally, the implementation of confidence indicators and feedback mechanisms will provide pilots with insights into the AI's decision-making processes, enhancing predictability and trust in the system.

Moving on to faulted conditions, a potential safety risk is the **pilot's lack of preparedness to handle unexpected situations** during system failures. To address this, extensive scenario-based training should be conducted to simulate abnormal conditions and system failures. Developing comprehensive emergency procedures and conducting practice drills will ensure pilots can confidently respond to unforeseen events without relying solely on the intelligent assistant.

To ensure the system's accuracy and reliability, rigorous testing and validation of the intelligent assistant are essential. Proper training and calibration of the AI model will contribute to minimising inaccuracies, enhancing the system's overall safety. Continuous monitoring and improvement of the system's performance are vital to its safe and effective operation in real-world flight scenarios. Safety considerations should remain a priority throughout the system's development and deployment stages.

3.4. Security assessment

The security assessment for the intelligent assistant system is currently in progress, with some aspects already identified and others yet to be fully evaluated.

The assessment indicates that **data is the primary asset together with the service providing the data, and supporting technology such as Bluetooth or Wi-Fi plays a secondary role**. Data is the foundation of the intelligent assistant system's decision-making, situational awareness, and overall performance. On the other hand, Wi-Fi, while essential for data transmission and communication, is considered a secondary asset, given its supporting role and potential limitations in availability.

Specific potential forms of attacks to which the intelligent assistant system could be vulnerable have not been fully defined yet, however the integrity of data provision could potentially be jeopardised by radio frequency (RF) attacks, specifically those of the jamming nature and it would be interesting to explore if such protocols are sufficiently secure for IA application. While the system is not considered a primary target for attacks, it is essential to consider potential threats to its security. Various attack vectors, such as denial-of-service (DoS) attacks, man-in-the-middle attacks, and data manipulation attempts, need to be explored to ensure robust protection against potential attacks.

The assessment acknowledges that the final decision-making authority lies with the pilot (even if startled), reducing the likelihood of significant adverse effects due to the intelligent assistant's operation. However, potential security breaches or failures could still have adverse consequences, even if the pilot retains ultimate decision-making power. A security breach in the IA could lead to unauthorised access or manipulation of critical flight data or expose the physiological data of the pilots. Similarly, a failure or malfunction within the IA's operational mechanisms could result in inaccurate information being presented to the pilot or the IA being unable to respond appropriately during

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

critical moments. Even with the pilot retaining decision-making authority, these potential adverse consequences cannot be overlooked. The IA's actions and recommendations are likely to carry significant weight, and if inaccurate or compromised, they could impact the pilot's ability to make informed decisions, especially during high-stress situations.

The evaluation of the intelligent assistant system's resilience against adversarial attacks or manipulation attempts is yet to be conducted. Understanding the system's vulnerabilities and the potential consequences of attacks is critical to implementing appropriate security controls. Robust authentication and access control mechanisms are essential to prevent unauthorised access and manipulation of the system's functionalities.

A proactive approach to security, continuous monitoring, and improvement will ensure the system's safety, integrity, and effectiveness in supporting pilots during flight operations. Security considerations should remain a priority throughout the system's development, deployment, and operation stages.

3.5. Liability assessment

3.5.1. UC1 legal framework

According to the results obtained in the previous assessment, the liability assessment of the UC1 is based on a limited set of information and aims to provide insights for the future development of the UC and the IAs. The contents reflect the state of the arts and provide only a prognostic high-level analysis.

From the legal perspective, UC1 presents some critical aspects. Notwithstanding cockpit operations have a well-established legal and regulatory framework, the introduction of the IA for startle response presumes the delegation of some tasks that fall within the pilot’s responsibilities.

This is the reason why the UC1 legal framework mainly focuses on the outline of the specific liability regime of the Pilot In Command (PIC), as the primary intended user of the IA. The considerations about the PIC are complementary liability regimes of the air carrier, as recruiter and employer. Eventually, the results obtained for these two categories of final users need to be integrated with the analysis of the liability regime of the manufacturer that contributes to the development and deployment of the new technology.

In particular, for the purposes of the UC1, the PIC is responsible for ensuring that a flight is not commenced or continued beyond if any flight crew member is incapacitated from performing duties by any cause such as injury, sickness, fatigue, the effects of any psychoactive substance, lack of oxygen (ICAO, Annex 6(II), § 2.2.5). The tasks of the PIC can be assigned to another flight crew member during the cruise, to allow the pilot-in-command or a co-pilot to obtain planned rest. In case of emergency, the PIC also needs to be able to execute procedures for crew incapacitation and crew coordination including allocation of pilot tasks, crew cooperation and use of checklists (ICAO, Annex 1, ibidem).

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

Delegation of function in case of incapacitation, therefore, is an integral part of flight crew training and does not represent a new feature. The novelty, however, is the addressee of this delegation of functions, i.e., no longer a trained member of the crew but an IA for startle response.

It is worth to be noted that, generally, nothing shall relieve the pilot-in-command of an aircraft from the responsibility of taking such action, including collision avoidance manoeuvres based on resolution advisories provided by the equipment (ICAO, Annex 2, § 3.2). Nonetheless, the applicability of this liability regime to PIC is conditioned by the correspondence of tools available to the information provided about the 'philosophy of the systems and the expectations of the final users (BFU, 2004).

3.5.2. Actor-based liability analysis

In light of the above, the PIC liability analysis relies on some basic assumptions. From the legal perspective, we are evidently before a **well-defined professional outline, with specific accountability positions and task responsibilities defined according to detailed duties defined by aviation law and regulation**. On these grounds, the civil liability regime of this actor should be related to the contractual relationship between employer and employee, coupled with the professional insurance required by law. The criminal liability outline, instead, would be deeply impacted by the accountability duties of this category of actors. Task responsibilities, therefore, should be considered beyond their nominal value and generally extended to the entire procedures, considered as a whole. This would include a general supervisory duty on the appropriate performance of other subjects' tasks (e.g., flight-crew members, if any) and on the functioning and reliability of the instrumentation available on the aircraft.

Against this background, a liability hypothesis for the pilot may be confirmed if the following conditions are jointly met:

- there is an **injury** to a legally protected interest;
- there is **careless behaviour** of the pilot;
- there is a **causal correlation** between the behaviour and the injury.

Some exceptions or counter arguments may be advanced, for instance, in case the pilot's behaviour lacked will.

However, it is essential to underline that the **PIC – as an end-user of the IA – needs to be enabled to properly perform her/his tasks with the support of the new tool**. Therefore, the actor-based liability hypothesis will be confirmed only if the causal link between the PIC's conduct and the injury is fully attributable to this actor.

If the negative occurrence is also correlated to other factors, not connected to the PIC's behaviour only, there could be secondary liability hypotheses. Producers may be involved if the injury is due to product defects affecting the functioning and usage of the IA. If instead the problems are due to poor implementation plans (e.g., due to the poor quality of the products and procedures, as well as of implementation and investment strategies), there might be an organisation liability hypothesis.

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

Considering the evolutive nature and the possibility of customisation of some AI-enabled technologies over time, these different liability hypotheses for developers and producers and implementing organisations can coexist and have contributory nature.

3.5.3. Liability assessment results for the UC1

In the use of the IA for startle response, the liability analysis discloses a deep intrinsic correlation among producers, employer organisations and the PIC strategies and behaviours.

In particular, **the organisations (e.g., airlines and air carriers) are responsible for all the organisational aspects of these innovations**. They are indirectly responsible for the behaviour of their employees in their interactions with the new tools and they have to ensure appropriate training and adequate usage conditions. Moreover, they are responsible for all the organisational aspects of these innovations, choosing only those products or solutions adequate for their operative purposes and tasks and reviewing all the internal policies and procedures impacted by the innovation.

Against this background, PIC should take into consideration all the legal risks associated with the use of the tool. On the one hand, there are **issues related to the overconfidence and over-reliance on the support of the IA in recovering from the startle effects**, and so an undue delegation of tasks for protective reasons. On the other hand, there may be **problems related to PIC's overconfidence in her/his recovering capabilities**, and so a premature resume of controls while still in a state of incapacitation. In case of accident, both these scenarios may embed the risk of professional negligence, due to careless actions and/or careless omissions. The only exclude should be the lack of will be due to the severe incapacitation of the PIC. This exception could exclude the liability of PIC but may not be sufficient to exclude the liability of other actors involved (air carrier, manufacturers).

3.6. Recommendations for the UC1

Looking at the future development of the UC1, these are the main recommendations addressed to UC1 owners to mitigate the possible risks emerged for the preliminary assessment:

- **To define qualitative and quantitative indicators about the detection and the end of the startle effects**. This could facilitate a more responsible use of the tools, improving the human self-perception and her/his situation awareness. This may also benefit the liability apportionment.
- **To avoid the use of physiological data for any purposes not strictly related to the smooth functioning of the tool**. This data can be qualified as sensitive data, since the information is used to detect and manage a situation of human vulnerability. More specifically, it should be granted this data will never be used for profiling and/or performance assessment.
- **To define adequate training for end users, ensuring they are well aware of the philosophy of the system as well as of its intrinsic limits**. The upskilling/reskilling process

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

should include ethics-based aspects related to the autonomy of the human agent, and so grant the PIC will be anyway able to manage the situation by her/his own, also following alternative procedures. The objectives and features of this training should be defined in parallel with the development of the IA, and progressively adapted according to its evolutions.

4. UC2 – IA for the flight deck route planning/ replanning

4.1. Concept description and possible scenarios

The UC2 works on the development of an IA for the flight deck route planning and replanning named **COMBI (i.e., Bidirectional Communicator), that aims to reduce the complexity of mission management. The intended users are pilots in commercial aviation**.

The IA will be used when an unexpected but non-critical event occurs (last-minute flight plan change, medical problem, airport closure, weather disruption, ...). It can also be used in a routine situation to improve situational awareness. Generally, the IA leverages many data sources, like those of the environment (weather, cloud cover, air humidity, temperature, etc.), air traffic data and airport conditions, but also information on the crew (fatigue, stress), passengers (type of journey, nationality) and cargo.

The intelligent assistant leverages the COMBI concept to enable shared understanding and joint resolution of complex situations. **By empowering pilots with advanced cognitive assistance, the system aims to improve decision-making quality and speed, enhance mission efficiency, increase success rates in achieving goals, reduce in-flight incidents, and ultimately simplify mission management through dynamic communication between users and the AI system**.

This interaction aims to optimise information exchange, enhance decision-making efficiency, and effectively reduce pilots' workload. By providing real-time updates, relevant data analysis, and proactive recommendations, the intelligent assistant becomes a co-pilot, facilitating collaborative decision-making and contributing to safer, more efficient missions.

In light of the above, the AI assistant is asked to perform different roles. Indeed, according to the Human-AI Teaming Types & Digital Assistants categories developed HAIKU WP3 (Table 1) the tool should act as:

- informer,
- secretary.

When acting as an **informer**, it provides information proactively. Instead, when it collaborates supporting decision-making on demand, the IA could also be considered a **secretary**.

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

4.2. HF assessment

The HF Assessment for the IA system in UC2 focuses on developing an intelligent assistant to reduce the complexity of mission management for commercial aviation pilots. The collaboration between pilots and the Intelligent assistant is characterised by moderate engagement, with ongoing efforts to define clearer collaboration goals.

The collaboration goals encompass both physical and knowledge-oriented aspects, emphasising understanding, decision-making, and action-taking. While emotional empathy is not involved, there is cognitive shared mental model empathy between humans and the AI, contributing to a proactive interaction pattern.

The Intelligent assistant system has limited agency, handling supervised tasks, while pilots retain full decision-making authority. The collaboration mainly occurs in a co-located physical context involving ATC personnel and the operational control centre.

Pilots are fully aware of and provide consent before interacting with the Intelligent assistant system. Presently, the consequences of system failure are low from a safety perspective. However, as AI systems become critical in complex situations, the potential consequences of failures may increase. Rigorous testing, validation, and fail-safe mechanisms are essential mitigations.

The benefits of the IA system are significant, including reduced stress and workload for pilots, improved situation awareness and decision-making, cost reduction, and enhanced safety in operations.

The Intelligent assistant system operates through a screen-based mode of interaction. Its adaptability is proactive in collaborative situations and reactive in cooperative scenarios.

A potential issue relates to the lack of communication of confidence levels by the IA system to humans. Transparent feedback on the AI's recommendations or decisions is crucial to prevent uncertainty and over-reliance on the system.

Another potential issue is the need to consider cultural and cognitive abstraction differences among pilots collaborating with the system. The interface and communication should be designed to accommodate diverse backgrounds, ensuring effective interaction.

Long-term adaptability is essential as the AI system becomes more critical in-flight operations. Developing a system that can learn from interactions and adapt to changing aviation practices and regulations is crucial.

Effective collaboration between pilots and the intelligent assistant requires seamless integration of human decision-making and AI-supported decision-making. To facilitate this, training programs should emphasise human-AI teaming, clarifying the roles and responsibilities of both pilots and the AI system. Encouraging open communication and a shared mental model between pilots and the AI will promote efficient decision-making.

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

Overall, the collaboration between pilots and the Intelligent assistant has the potential to optimise flight operations, leading to enhanced safety, efficiency, and decision-making. Addressing potential issues and implementing appropriate mitigations will contribute to the success of this intelligent assistant in commercial aviation.

4.3. Safety assessment

During the safety assessment of the use case in its initial design phase, many questions remained unanswered due to the early stage of development. Detailed risk-related aspects and safety measures have yet to be fully defined.

The primary focus of the initial design analysis under normal operations is to identify risks, risk metrics, and risk levels specific to the use case. This information is crucial in formulating clear risk mitigation strategies to address identified safety risks. One critical safety concern pertains to the **accuracy and reliability of the input data** provided to the intelligent assistant system. Inaccuracies or inconsistencies in the data could lead to incorrect decisions and compromised safety. To mitigate this data-related risk, robust data quality assurance processes should be implemented. Continuous assessment of data integrity through data validation, error-checking algorithms, and real-time data monitoring will help identify and promptly address any data anomalies.

Potential over-reliance on the intelligent assistant system by pilots may lead to complacency and reduced situational awareness. To prevent this, comprehensive training on the proper use of the system and awareness of its limitations are necessary. Emphasising the importance of cross-verifying critical information and maintaining situational awareness will help pilots avoid over-reliance on the AI system.

Additionally, the safety assessment considers potential **risks associated with abnormal conditions**. Acknowledging the pilot's ability to modify parameters or input data to match their mental model is crucial. To ensure safety, the intelligent assistant must be evaluated for robustness and reliability under different operating conditions and potential failure scenarios. Implementing failsafe fallback plans will effectively address errors or faults in the system. Mitigating the risk of abnormal data manipulation by pilots involves implementing data integrity checks and monitoring for unauthorised modifications. Maintaining a clear audit trail of data changes aids in post-incident analysis and accountability.

By addressing these potential safety issues and implementing appropriate mitigations, the development and deployment of the intelligent assistant in commercial aviation can maintain a high standard of safety and reliability, ensuring effective support to pilots in mission management and decision-making processes.

4.4. Security assessment

UC2 is still in the initial design phase, and many of the detailed security-related aspects and measures have

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

not been fully defined yet. The security-related considerations and assessments are expected to be addressed in subsequent stages of development and deployment.

Inferences can be made at this stage. Critical assets at risk include flight operations data, passenger and crew information, and the communication infrastructure, which require protection to ensure safe and efficient flight operations.

One of the primary concerns is **unauthorised access**, where individuals gaining entry to the intelligent assistant system could compromise its functionality and integrity, potentially leading to safety risks and disruptions in flight operations. Additionally, the possibility of a data breach raises concerns about the privacy and security of sensitive information, as unauthorised exposure could result in misuse and privacy violations.

The system may also be **susceptible to malware and cyber-attacks**, which could disrupt operations, compromise functionality, and introduce safety risks. Moreover, potential DoS attacks may overwhelm the system, causing service disruptions and hindering mission management capabilities.

To address these potential security issues and vulnerabilities, several security controls and mitigations should be implemented. Robust authentication mechanisms, including multi-factor authentication, should be employed to prevent unauthorised access. Data encryption at rest and during transmission can protect sensitive information from unauthorised access. Regular assessments and updates of software, along with code reviews and security testing, can mitigate software vulnerabilities.

4.5. Liability assessment

4.5.1. Legal considerations about the UC2

According to the results obtained in the previous assessments, the liability assessment of the UC2 is based on a limited set of information and aims to provide insights for the future development of the UC and the IAs. The contents reflect the state of the arts and provide only a prognostic high-level analysis.

Since this UC focuses on cockpit operations and the primary users are pilots, the legal considerations about this scenario are similar to the ones presented for the UC1.

4.5.2. Actor-based liability analysis

Since this UC focuses on cockpit operations and the primary users are pilots, the actor-based liability analysis about this scenario is similar to the one performed for the UC1.

4.5.3. Liability assessment results for the UC2

As per the UC1, the liability analysis discloses a deep intrinsic correlation among producers, employer organisations and the PIC strategies and behaviours.

In particular, **the organisations (e.g., airlines and air carriers) are responsible for all the organisational aspects of these innovations**. They are indirectly responsible for the behaviour of their

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

employees in their interactions with the new tools and they have to ensure appropriate training and adequate usage conditions. Moreover, they are responsible for all the organisational aspects of these innovations, choosing only those products or solutions adequate for their operative purposes and tasks and reviewing all the internal policies and procedures impacted by the innovation.

Considering the liability regime applicable to the PIC and the scenarios prospected in UC2, there are issues related to overconfidence and over-reliance on the support of the IA. In case of accident, these scenarios may embed the risk of professional negligence, due to careless actions and/or careless omissions.

However, it is interesting to note how the results of the liability assessment for the UC2 differ from the ones obtained for the UC1, and this despite the role of the human agent involved. More specifically, in UC2 there are limited risks related to the possible incapacitation of pilots. Therefore, **the main liability concerns regard the issues associated with overconfidence and over-reliance in the use of the IA, as well as on the capabilities to understand the ordinary limits of the adopted technologies and to manage the same tasks with and without the support of this latter**. Anyway, comparing UC1 and UC2, if the conditions that may form a liability hypothesis are the same, the associated risk is substantially different.

4.6. Recommendations for the UC2

Looking at the future development of the UC2, these are the main recommendations addressed to UC2 owners to mitigate the possible risks emerged for the preliminary assessment:

- **To always ensure an adequate level of situation awareness**. This could facilitate a more responsible use of the tools and may also benefit the liability apportionment.
- **To define adequate training for end users, ensuring they are well aware of the philosophy of the system as well as of its intrinsic limits**. The upskilling/reskilling process should include ethics-based aspects related to the autonomy of the human agent and so grant the PIC will be anyway able to manage the situation by her/his own, also following alternative procedures. More specifically, in UC2, it is advisable to pay particular attention to direct and indirect issues concerning human autonomy and dignity (free self-determination in decision-making) accountability (of decision-making and its consequences) fairness (accessibility and universal design, now and over time, also in light of the background of users) and societal well-being impact on work and skills. The objectives and features of training should be defined in parallel with the development of the IA, and progressively adapted according to its evolutions.

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

5. UC3 – IA for UAMC to assist in traffic management

5.1. Concept description and possible scenarios

The UC3 works toward the development of an **IA aimed to support the safe delivery of U-Space Services (USSs)** in the time window 2030-2050.

The primary user of the tool is the **Urban Air Mobility Coordinator (UAMC), which will be helped by the IA in managing her/his U-Space (US) area**. As a result, the system is generally indicated as Digital assistant for UAM Coordinator (DUC).

As outlined in D3.2, the UC3 potentially embraces several key USSs in *Z volume,* including network identification; geo-awareness; flight authorisation; traffic information; weather information, and conformance monitoring. In this scenario, the UAMC will have a key human role as part of U-Space Traffic Management (UTM) for a specific city, providing real-time strategic and tactical U-space services to UAS and UAM operators and stakeholders. Performing her/his tasks, the UAMC has to manage large traffic volumes safely and efficiently in cities. Therefore, from an operative point of view, this actor reasonably needs to gather and provide the most updated information/data as well as to ensure information/data validity, to successfully communicate and coordinate with all airspace users/operators and UAM stakeholders and to effectively support and handle emergencies and contingencies. Performing these tasks, the UAM Coordinator will be supported by intelligent assistants capable of monitoring all traffic in the city airspace as well as monitoring ground events and city life with an impact on trajectory planning.

More specifically, **the DUC will care for the majority of standard, repetitive, normal tasks (e.g., flight authorisation, traffic monitoring, flight information, and weather information). So doing, the DUC allows the UAM Coordinator to focus on high-level strategic decision-making** in oversight of UAM operations, reducing human task-/workload. DUC will support the UAM Coordinator in day-to-day normal operations and emergency situations such as in-flight medical emergencies.

In this regard, the DUC would be active 24/7, constantly elaborating data coming from the CIS and other relevant sources/systems, e.g., USSP and UAS/UAM operators. The data include but are not limited to airspace, weather, population density, historical data and data related to city life (such as road traffic, events, emergencies etc.). In principle, the DUC could be operated using a tablet or a large touch-based Human-Machine Interface (HMI) that the UAM Coordinator can interact with via mouse and keyboard, or touchscreen. Moreover, by means of visual and auditory stimuli, the DUC could produce visualisations of real-time data facilitating information viewing/inspection and decision-making process and attracting the UAM Coordinator's attention to critical situations. The UAMC should be also able to input/insert data and steer the intelligent assistant (e.g., by adjusting higher-level parameters according to key performance) by means of textual interactions.

At this stage of the design process, the IA would be responsible for the following tasks (tentatively):

- gathering and exchanging data from the CIS and other relevant sources/systems,

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

- integrating real-time information from city ground events/social life of the city and airspace activities,
- executing day-to-day repetitive tasks e.g., flight authorisation, traffic planning/monitoring, identifying needs for geo-fences in response to situations on the ground and giving clearances etc.,
- identifying violations and detecting deviations from the norm,
- directing the UAM Coordinator's attention to any kind of abnormal situations/events,
- supporting contingencies and emergencies e.g., determining the quickest trajectory to get from A to B, dynamically establishing priority criteria, and coordinating with actors concerned,
- Provide an explanation for why a certain output (e.g., suggested route) is motivated,
- relaying information from one actor to another.

The UTM CITY interface offers a comprehensive view of UAS/UAM traffic, services, and airspace restrictions on a city map with a dashboard. Through this interface, the UAM Coordinator can communicate with various UAM stakeholders, including UAS/UAM operators, vertiports, logistics hubs, emergency stakeholders like hospitals, and other traffic management services. The interface serves as the primary point of interaction with the DUC as well. Typically, the UAM Coordinator does not engage in detailed monitoring or interaction with individual flights. Instead, they work at a higher level of abstraction, determining traffic separation objectives that the DUC implements or establishing priorities, especially in emergencies or problematic situations. The DUC communicates with the UAM Coordinator directly through the interface, providing attention guidance, suggestions, feedback, and explanations via a dedicated DUC Communication window. The UAM Coordinator interacts with the DUC through this window, seeking explanations, assigning tasks, adjusting performance parameters, and modifying the level of automation of the DUC as necessary.
In light of the above, the AI assistant is asked to perform different roles. Indeed, according to the Human-AI Teaming Types & Digital Assistants categories developed HAIKU WP3 (Table 1) the tool should act as:

- coordinator;
- executor;
- observer;
- informer.

As a **coordinator**, at a higher level of cognitive control, the DUC is expected to coordinate the operations in the city sky. As an **executor** (albeit at the lower levels of cognitive control), the DUC is expected to automatically perform low-level control and communications as well as repetitive tasks. Eventually, as an **observer** and **informer**, the DUC is expected to perform some observation and information tasks to establish compatible and shared situation awareness between the DUC and the UAMC.

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

5.2. HF assessment

The HF Assessment aims to provide a comprehensive understanding of the cooperation between the UAM Coordinator and the Intelligent assistant, as well as the characteristics of the AI system. It emphasises the importance of shared goals, alignment, and continuous cooperation to ensure the effective performance of the Intelligent assistant in supporting UAM operations safely and efficiently. The assessment also delves into the potential impact of the human-AI cooperation on decision-making, information processing, and traffic coordination.

Currently, the Intelligent assistant is evolving over time through model updates and continual interaction, rather than being fixed once deployed. This ongoing cooperation between the Intelligent assistant's developers and the system plays a crucial role in refining its capabilities and ensuring optimal performance.

The goals of the **human-AI cooperation are generally clear, focusing on reducing the UAM Coordinator's workload and maintaining the safety and flow of UAM operations**. This cooperation involves both intellectual and motivational aspects, fostering shared control between the UAM Coordinator and the Intelligent assistant.

However, there are potential issues in the cooperation that need to be addressed. One potential issue is human-AI decision alignment. **Divergent interpretations of goals or priorities could lead to conflicting decisions. Transparency in the decision-making process could foster mutual understanding**.

Overreliance on AI is a potential issue. The UAM Coordinator might overly depend on the Intelligent assistant, leading to complacency or reduced situational awareness. Maintaining a balance of responsibilities and continuous training for the UAM Coordinator can help prevent overreliance.

The cooperation between the UAM Coordinator and the Intelligent assistant is continuous and repetitive, persisting as long as UAM operations are ongoing. It takes a mixed hybrid approach, combining concurrent cooperation with occasional turn-taking for specific situations or occurrences. **Turn-taking challenges during specific situations might cause miscommunication or delays in handing over control. Designing the interface for seamless transitions and conducting simulations and training scenarios can help address this issue**.

The primary mode of interaction between the UAM Coordinator and the Intelligent assistant is via a screen-based interface, with potential consideration for voice-based interactions in certain scenarios.

The system's predictability is high, ensuring consistent and reliable performance. The occurrence of false positives and false negatives by the system is yet to be evaluated. While the system's ability to communicate confidence levels to the human is not a primary focus at this stage, it may be considered in future research. The specifics of how the Intelligent assistant communicates its decision-making process to the human are part of ongoing research objectives.

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

Efforts are being made to avoid any human-like characteristics in the Intelligent assistant to prevent anthropomorphism and maintain a clear distinction between the roles of human and AI.

The primary users cooperating with the Intelligent assistant are adults within the age range of 18-65. These users are highly trained and possess significant previous technology interaction experience. The assessment also acknowledges the potential presence of third parties in the cooperation, the co-location of human and Intelligent assistant agents, and the importance of a high safety culture environment in this context.

Clear communication, shared goals, ongoing feedback, and appropriate training are essential to overcome challenges and achieve successful decision-making and coordination in the human-AI cooperation for UAM operations.

5.3. Safety assessment

At this stage of the design process, a comprehensive safety assessment for the intelligent assistant system was not possible. While the initial design analysis under normal operations and abnormal conditions involved identifying risks, risk metrics, risk levels, and potential consequences, the implementation of clear risk mitigation strategies and safety measures for continuous assessment of data quality and system accuracy was still pending. Additionally, the evaluation of the intelligent assistant system's robustness and reliability under different operating conditions and potential failure scenarios, along with the development of mechanisms to trigger new safety reviews when changes occur, were yet to be established. Similarly, the implementation of failsafe fallback plans to address errors and low-confidence results was not in place. Given the ongoing nature of the development and the need for further testing and validation, a comprehensive safety assessment will be conducted in subsequent stages.

At the moment the potential safety issues and mitigations could be the following. The **system's inability to communicate confidence levels to the human operator** could lead to uncertainty and potential errors in decision-making. To mitigate this, efforts should be made to enhance the system's ability to communicate confidence levels and uncertainties to the UAM Coordinator, enabling informed decision-making based on the system's reliability.

The UAM Coordinator's **excessive dependence on the AI system** could lead to reduced situational awareness and complacency. To address this, a balanced approach to collaboration between the human and the AI is necessary. The UAM Coordinator should remain proficient in UAM operations and maintain a high level of situational awareness, using the intelligent assistant as a supportive tool rather than a substitute for critical decision-making.

5.4. Security assessment

At the current stage of development, a comprehensive security assessment for the intelligent assistant system was not conducted. The detailed analysis and definition of potential forms of attacks and their

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

adversarial, critical, or damaging effects were yet to be addressed. Similarly, the extent of the intelligent assistant system's exposure to cyber-attacks and its impact on rights such as privacy, physical and mental integrity, and data protection were still pending. Furthermore, the identification of controls to ensure the system's security was not completed.

In this context, the primary asset is the DUC system which serves as the pivotal hub for orchestrating and managing U-Space activities. The secondary assets encompass critical components that sustain the primary asset's functionality. These include the communication infrastructure, data integration sources, operational algorithms, user interaction interfaces, and the DUC's software and hardware components. In the use case involving the intelligent assistant system for UAM operations, one significant security concern could be the potential for unauthorised access to sensitive data and system control, which could lead to malicious attacks and compromise the safety of UAM operations.

Potential forms of attack include **unauthorised access, data interception, DoS attacks, and malware injection**. Unauthorised access could allow malicious actors to manipulate UAM operations data or disrupt traffic coordination. Data interception might compromise communication between the UAM Coordinator and the intelligent assistant, leading to the misuse of sensitive information. DoS attacks could overwhelm the intelligent assistant system, causing service disruptions and hindering effective UAM management. Malware injection might result in unauthorised data access, corruption, or system manipulation.

To address these security risks, several mitigation strategies should be implemented. Strong authentication, such as multi-factor authentication (MFA), can ensure that only authorised users can access the intelligent assistant system. Encryption and secure communication protocols should be in place to prevent data interception. Strict access controls and role-based privileges can limit access to necessary functions and data, and regular reviews should be conducted to prevent unauthorised access.

Intrusion detection and prevention systems (IDPS) can monitor and detect suspicious activities, triggering immediate responses to block potential attacks. Regular security audits and penetration testing are essential to identify vulnerabilities and address them promptly. Redundancy and failover mechanisms can ensure continuous availability of critical UAM operations and mitigate the impact of DoS attacks.

5.5. Liability assessment

5.5.1. Legal considerations about the UC3

The liability assessment of the UC3 is based on a limited set of information and aims to provide insights for the future development of the UC and the IAs. The contents reflect the state of the arts and provide only a prognostic high-level analysis.

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

From the legal perspective, this UC presents several critical aspects. **UASs are a relatively recent phenomenon in civil aviation, and thus the regulation of these new technologies currently represents a challenge for aviation law**.

As highlighted by literature (Bauranov & Rakas , 2021) (Cohen, Shaheen, & Farrar, 2021) (Fiallos Pazmiño, 2020) (Scott, 2022). The rules and standards usually applicable to air traffic management (ATM) have primarily been designed to manage manned air traffic and they are not easily transposable to unmanned systems, mainly due to the differences among the two. Moreover, the ATM already handles an enormous amount of air traffic and is reaching its maximum capacity.

This is the reason why the EU still lacks a specific liability regime for USSPs and their operators. Therefore, for the purposes of the liability assessment, we consider two different outlines for the liability profile of the UAMC and its USSP. On the one hand, according to the present legal framework where we have no specific reference to this actor and its related liability regime, we will analyse the liability risks exposure of the UAMC taking into account only the conditions that may generate general negligence. On the other hand, instead, having in mind the analogies between ANSPs and USSPs, we will approach the same situation trying to qualify the UAMC according to the ATCO liability regime. It is essential to bear in mind that this last configuration does not reflect the current state of the art of EU law but may provide useful insights for a proactive approach to liability risks from a medium/long-time perspective. The results obtained by this first round of the liability assessment have to be considered temporary and may be reviewed or updated in the future releases of D7.3, also in light of the evolution of the regulatory frame of reference.

5.5.2. Actor-based liability analysis

From the analysis proposed in section (UC1) it is possible to derive a high-level view of possible liability issues that derive from the potential use of the UC3 solution. In particular, this section of the report will identify the possible variations of liability regimes that derive from the use of DUC for the various actors involved, taking into consideration the preliminary list of tasks assigned to the UAMC and the DUC.

5.5.2.1. The UAMC, as a new generic actor

In light of the above, the first hypothesis to analyse considers the **UAMC as a new actor with a non-specific liability regime**. This situation more faithfully represents the current legal framework but takes no consideration of the imminent evolutions of this latter. The applied standards thus refer to actions and/or omissions that may raise generic negligence for professional actors.

From a liability perspective, therefore, we are before a generally flexible professional outline, with a non-specific accountability position and task responsibilities defined according to her/his specific duties. On these grounds, this actor should be subject to a civil liability regime defined by her/his employment contract and covered by vicarious liability from her/his. From the criminal liability perspective, instead, s/he should be subject to general negligence standards and assessed according to a professional duty of care.

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

In light of the above, in the transition towards a more comprehensive legal framework for the UASs liability regime the most immediately applicable legal regime for the UAMC should be generic negligence, as defined for aviation professionals. A liability hypothesis can be confirmed if the following conditions are jointly satisfied:

- there is an **injury** to a legally protected interest;
- there is **careless behaviour** of the person at stake[1]; and
- there is a **(causal) relation** between the behaviour and the injury.

Some exceptions or counter arguments may be advanced, e.g., the fact that the person's behaviour lacked the will.

In light of the above, the UAMC has no well-established accountability position. This is why it is reasonable to assume that**, if qualified in these terms, the UAMC is responsible for her/his own tasks but has context-limited proactive duties related to the procedures performed by actors not directly instructed by him/her**.

However, it is essential to underline that the **UAMC – as an end-user of the IA – needs to be enabled to properly perform her/his tasks with the support of the new tool**. Therefore, the actor-based liability hypothesis will be confirmed only if the causal link between the UAMC's conduct, and the injury is fully attributable to this actor.

If the negative occurrence is also correlated to other factors not only correlated to the UAMC's behaviour, but there could also be a secondary liability hypothesis. These will be on the shoulder of producers if the injury is due to product defects affecting the IA. If instead the problems were due to poor implementation plans (e.g., the quality of the products and procedures, as well as implementation and investment strategies), there might be an organisation liability hypothesis. Considering the evolutive nature and the possibility of customisation of some AI-enabled technologies over time, these different liability hypotheses for developers and producers and implementing organisations can coexist and have contributory nature.

5.5.2.2. The UAMC, as an ATCO equivalent figure

On the other hand, **the UAMC could be equated and trained as ATCOs**. This choice would rely on the proximity of their respective functions and task responsibilities. In this case, the civil liability regime should be related to the contractual relationship between employer and employee, coupled with the

[1] Careless behaviour may consist of a careless action or a careless omission. An individual's behaviour is careless when the person took action, and the action was careless. Carelessness is usually determined by assessing whether the action violates the standard of due care, which is the proper behaviour that a professional operator would have been required to follow in the given situation. Such expectations depend on the tasks assigned to the UAMC, as well as on international and national laws, public or private standards and regulations, or even customs.
An individual's omission will be careless when the person failed to take action; the person had a duty to act; and the person's action would have prevented the injury. The content of the duty to act will depend on the tasks assigned to the UAMC, as well as on international and national laws, public or private standards and regulations, or even customs.

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

professional insurance required by law. The criminal liability outline, instead, would be deeply impacted by the accountability duties of this category of actors. Task responsibilities should be considered beyond their nominal value and generally extended to the entire procedures, considered as a whole. This would include a general supervisory duty on the appropriate performance of other subjects' tasks (e.g., remote pilots, and other USSPs).

In light of the above, if equated to an ATCO, a liability hypothesis for the UAMC can be confirmed if the following conditions are jointly satisfied:

- there is an **injury** to a legally protected interest;
- there is **careless behaviour** of the person at stake; and
- there is a **(causal) relation** between the behaviour and the injury.

At a glance, there are some similarities to the general negligence scheme. However, it is essential to note that the content of careless behaviour in this case is associate to **more severe accountability duties**.

The criteria previously described for assessing the lack of proper duty to care are set out with greater detail by international and national laws (such as navigation codes) public or private standards and regulations, or even customs. However, it is important to consider that ATCOs professional standards are carefully and systematically defined by the Air Traffic Control Procedures Manual; a reference evidently not applicable to UACM per se. In this regard, a possible exception may be based on the argument that the particular action omitted by the ATCO is not imposed in the future USSPs Procedures Manual. In fact, in several legal cases, the controller who has fulfilled his obligations (as imposed in the manual) has been cleared of further liability.

Hence, **if qualified as a simile-ACTO, the UACM may incur a sui generis accountability position. Beyond her/his specific tasks and the related responsibilities, in this case, s/he may also be considered accountable for the tasks performed by other operators and subject to a general duty of care and proactive attitude to monitor and prevent potential risk situations**.

In principle, both these options are plausible per se. However, the choice between these two should take into account the consequences of the liability risks exposure of the other subjects involved, especially the ATCOs.

Even in this case, it is essential to underline that the **UAMC – as an end-user of the IA – needs to be enabled to properly perform her/his tasks with the support of the new tool**. Therefore, the actor-based liability hypothesis will be confirmed only if the causal link between the UAMC's conduct, and the injury is fully attributable to this actor.

Liability hypothesis for developers and producers and implementing organisations can be confirmed as explained above, as well as coexist and have contributory nature according to specificities of the different operative scenarios.

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

5.5.3. Liability assessment results for the UC3

In light of the above, the liability analysis discloses a deep intrinsic correlation among producers, employer organisations and the UAMC strategies and behaviours.

In particular, **the organisations (e.g., USSPs) are responsible for all the organisational aspects of these innovations**. They are indirectly responsible for the behaviour of their employees in their interactions with the new tools and they have to ensure appropriate training and adequate usage conditions. Moreover, they are responsible for all the organisational aspects of these innovations, choosing only those products or solutions adequate for their operative purposes and tasks and reviewing all the internal policies and procedures impacted by the innovation.

Considering the tasks assigned to the **UAMC at the current stage of the HAIKU project, it is reasonable to presume this actor will have an ad hoc liability regime and specific accountability duties**. This is why the here reported observations will need to be further explored, also in light of the evolution of the current legal framework for UASs.

As mentioned, **generic professional negligence and ATCO negligence regimes do not differ for the conditions that may raise liability risks but for the accountability duties of the considered actor**. In this regard, the application of the Legal Case methodology highlights the following critical scenarios, mainly related to negligence due to careless actions and careless omissions.

Generally, we assume that in the performance of her/his duties, the UAMC always has a duty to act, especially if the action would prevent casualties, injuries and/or damages. In light of this, risks related to careless actions and/or omissions are mainly related to factual conditions in the cooperation between the UAMC and the DUC. Over-confident and/or over-reliant behaviour may expose the operator to the risk of careless actions. Scepticism and mistrust, as well as poor explainability and transparency features in the HIM, may lead to careless omission due to the underestimation of the suggestions provided and the tasks performed by the DUC.

5.6. Recommendations for the UC3

In light of the above, looking at the future development of the UC3, these are the main recommendations addressed to developers to mitigate the possible risks emerging from the preliminary analysis:

- **To develop a HMI sensitive to ethics-based aspects related to the autonomy of the human agent as well as explainability needs**. Where feasible, this care should take into account the operative needs related to situation awareness, better understanding of IA's decisions, capability of overriding these decisions, possibility to assess and integrate the data used for the IA's decisions. This is advisable to endure high level of compliance to direct and indirect requirements concerning: human autonomy and dignity (free self-determination in decision-making) accountability (of decision-making and its

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

consequences) fairness (accessibility and universal design, now and over time, also in light of the background of users) and societal well-being impact on work and skills.

- **To ensure the USSC an adequate explanation, interpretation and comprehension of the results provided by the DUC**. These aspects should be considered in light of human reaction time and her/his operative time window. The use of counterfactual evidence for rectifications could be useful. A clear definition of these factors may benefit the liability allocation among the actors involved.
- **To define adequate training for end users, ensuring they are well aware of the philosophy of the system as well as of its intrinsic limits**. The skilling process should include ethics-based aspects related to the autonomy of the human agent, and so grant the USSC will be able to manage the situation by her/his own, also following alternative procedures. The objectives and features of this training should be defined in parallel with the development of the IA, and progressively adapted according to its evolutions.
- **On a long-term perspective, it could be helpful to lobby for the introduction of clear conditions for defining and limiting the criminal liability of USSCs** when an incident/accident is due to the poor functioning of the DUC. In this regard, traceability will have a crucial value.

6. UC4 – IA for tower (and remote tower)

6.1. Concept description and possible scenarios

The UC4 focuses on the development of **ISA, an Intelligent Sequence Assistant, able to support the tower ATCO by suggesting a sequence for the aircraft using the runway in the most efficient way and maximising the number of aircraft using the runway in an hour**, while reducing stress from constant decision making and minimising potential errors/inefficiencies. In particular, this IA is primarily intended to support the arrangement of landing and take-of sequence. However, it may have a relevant role in the management of the runways and taxiway, contributing to tower control intended as a whole.

The primary users of this tool would be the ATCOs, who should work as a team with the assistant to safely maximise the use of the runway, especially during 'peak' hours. In this regard, the IA will help the ATCO make the best possible decisions. For the purposes of HAIKU, the UC4 is based in Alicante (Spain) and, prospectively, the target landscape are future digital towers of any airport where all the environment is processed and stored through cameras as digital data.

In this regard, the ISA would be active 24/7 to have the "bigger picture" of the surrounding operations environment, and not only of the most immediate upcoming events. The ATCO will have a visual representation of the suggested sequence, by means of a dedicated HMI. **If the ATC decides not to**

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

follow the suggested sequence, the assistant will recalculate it according to the new situation, and to everything that happens in a wider time window.

At the present (M12), the main benefits associated with the use of this IA concern the possibility to optimise the use of the runway, while reducing stress from constant decision making and minimising potential errors/inefficiencies. On the other hand, a potential pain point could be the time needed by ATCOs to understand and trust the suggested sequence.

In light of the above and according to the Human-AI Teaming Types & Digital Assistants categories developed HAIKU WP3 (Table 1) the tool should act as a **coordinator**. Indeed, the ATCO and the IA form a team that works together to safely maximise the use of the runway, especially during 'peak' hours. The IA will proactively help the ATC make the best possible decisions.

6.2. HF assessment

The HF analysis of the Intelligent Sequence Assistant (ISA) for air traffic control tower operations focused mainly on the cooperation between ATCOs and the ISA. The system's primary benefits include optimising runway utilisation and reducing ATCO stress, while potential challenges involve the human's trust and reliance on the ISA's suggestions. Overall, the successful implementation of the ISA promises to significantly enhance air traffic control operations and runway efficiency.

The HF analysis of the ISA highlights the importance of trust between ATCOs and the ISA. The system's proactive suggestions and predictable performance contribute to its potential benefits, such as optimising runway utilisation and reducing ATCO stress.

The ISA operates as an evolving Intelligent assistant, facilitating ongoing cooperation between the developers and the system. **While the system offers proactive suggestions, the final decision-making and action-taking remain within the purview of the human ATCOs**. ISA leverages a larger dataset and wider focus to provide optimal sequence suggestions.

The primary goal of the human-AI cooperation is to optimise runway utilisation, reducing the stress associated with constant decision-making while minimising potential inefficiencies. The ISA acts as a support system, enhancing ATCO performance by suggesting efficient aircraft sequences. Empathy is not a precondition for cooperation, and the human and the ISA have aligned goals with a shared understanding of the situation. In terms of agency, **the ISA does not play an active role in decision-making or action-taking, solely providing suggestions for ATCO consideration**. The human ATCOs retain full control and are responsible for finalising decisions based on the ISA's recommendations.

Three potentially complicated scenarios might result from the cooperation:

- **In the event of an ISA malfunction or failure, ATCOs must be prepared to take full control of the operations and manage air traffic safely without the AI's assistance**. Adequate training and clear standard operating procedures are essential in such situations.

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

- **If there is conflicting information, ATCOs should have the authority to prioritise their judgements** if ISA suggests a sequence conflicting with their observations or other systems. Open communication and the ability to override ISA's decisions are critical in such cases.
- **ATCOs might over-rely on ISA**. To mitigate this, regular training sessions should be conducted, emphasising the importance of maintaining situational awareness and independent decision-making among ATCOs. By understanding the limitations of the AI system and the need for human judgement, ATCOs can avoid blindly relying on the ISA's suggestions.

The ISA cooperation occurs in a co-located physical environment, with no involvement of third-party individuals or other Intelligent assistant systems. The human ATCOs are fully aware of the interaction with the ISA, and no explicit consent is required before engaging with the system. The potential consequences of the ISA's failure to perform as expected are considered moderate, mainly due to the possibility of blindly following the assistant's suggestions, which may lead to challenges for ATCOs. However, **the overall benefits, such as reducing ATCO efforts and optimising runway utilisation, are perceived as high**.

Regarding privacy and security concerns, typical users may have moderate levels of consideration, especially when it comes to the efficiency metrics employed by the ISA, which may raise some apprehension among ATCOs.

The interaction between the human and the ISA occurs through a screen interface, allowing the system to anticipate the next steps and proactively recalculate suggestions based on real-time data. The ISA's performance is predictable, provided the system's limits are understood. The system does not communicate its confidence levels or decision-making process to the human, presenting only the outcome of its calculations. **The assessment of the outcome of the cooperation is both subjective and objective, as it depends on objective metrics and on how ATCOs perceive and utilise the assistant**.

The user cooperating with the ISA is an adult ATCO with no special needs or accommodations. While there are no cultural norms specifically associated with the ISA, the user is familiar with working in similar systems. However, it is important to consider that a certain variability is expected in the traffic management style of the different prospective actors and this aspect shall be taken into account in the design of the ISA.

One crucial aspect is **building trust and reliance on the ISA's suggestions**. This can be achieved by designing a Human-Machine Interface (HMI) that provides clear explanations and justifications for the AI's recommendations. When ATCOs can comprehend the reasoning behind the suggestions, they are more likely to trust the system and work collaboratively with it.

In the event of a failure in the ISA's suggestions, thorough testing and simulations during the development phase become essential. By identifying and rectifying potential issues early on, ATCOs can effectively manage situations even without the AI's support. Having contingency plans and well-established procedures will further enhance their ability to handle unexpected situations.

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

Lastly, the limited explainability of ISA's decisions can be mitigated by implementing explainable AI methods. This will provide ATCOs with insights into the AI's decision-making process, making its recommendations more understandable and fostering better trust in the system's capabilities.

6.3. Safety assessment

The safety assessment encompasses an analysis of potential risks associated with the ISA. One critical risk that has been identified is related to the **system's suggestions potentially conflicting with the ATCO's risk perception**. The ISA, as a proactive suggester, may propose a sequence for aircraft movements that the ATCO deems too risky or challenging to handle, leading to potential safety concerns.

To address this risk, a prudent mitigation strategy involves adapting the system's suggestions to align more closely with the ATC's comfort level. By fine-tuning the assistant to be more conservative in its recommendations, it can better accommodate the ATCO's decision-making preferences and risk thresholds. This adaptation process will take into account the ATCO's expertise and experience, tailoring the system's suggestions to match their level of comfort and safety standards. This adaptability ensures that the cooperation between the ATC and the ISA is more effective and aligned with safety standards, making it a valuable tool in enhancing air traffic control operations. This approach not only helps prevent situations where the ATCO feels uncomfortable or overwhelmed with the system's suggestions but also ensures a smoother cooperation between the human and AI. The ATC retains full agency in the decision-making process while still benefiting from the ISA's support in optimising runway utilisation and reducing stress.

To ensure data quality, measures have been implemented to continuously assess the input data to the intelligent assistant system. This involves subjecting the system to traffic control simulations and obtaining valuable feedback from ATCOs, facilitating ongoing data quality assessment.

Moving on to the analysis under abnormal conditions, the safety assessment identifies the risk of **potential misuse or inappropriate use** of the intelligent assistant system. The issue revolves around the risk of overreliance on the ISA. To mitigate this, regular training sessions should be conducted for ATCOs, emphasising the importance of maintaining situational awareness and independent decision-making. ATCOs must understand the limitations of the AI system and recognize the necessity of their human judgement. By doing so, they can avoid blindly relying on the ISA's suggestions. Furthermore, simulating scenarios involving ISA malfunctions or inaccuracies during training prepares ATCOs to handle situations when the system is not available or functioning correctly. This ensures that they can make well-informed decisions independently, even without the assistance of the AI.

In worst-case scenarios, the system **might fail to recognize emergencies, leading to inappropriate suggestions** that could jeopardise air traffic control operations. To minimise this risk, thorough testing and simulations should be conducted during the system's development phase. Through comprehensive testing, potential issues can be identified and rectified, ensuring that the AI's suggestions align with real-world scenarios. Involving ATCOs in the testing process allows them to

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

provide valuable feedback and validate the system's performance in various operational conditions. Additionally, establishing clear procedures for ATCOs to manage situations effectively even without the AI's support is vital in maintaining safety during critical moments.

Regarding failsafe fallback plans, given the nature of the ISA as an addition to the existing control system, specific fallback plans are not deemed necessary. In the event of system failure, the ATCs will continue their role as controllers, maintaining control of the operations in the conventional manner.

As of the current status, the safety assessment is still pending the development of a mechanism for periodic evaluation of the system's robustness and reliability under different operating conditions and potential failure scenarios and the specification of procedures to handle cases where the intelligent assistant system yields results with low confidence scores.

6.4. Security assessment

The security analysis begins with the identification of primary assets that could be affected in the event of outages, attacks, misuse, or threats associated with the intelligent assistant. Data is recognized as the primary asset, vital for the efficient functioning of the system.

One critical potential issue that requires attention in the security assessment of the intelligent assistant system is the **vulnerability to various forms of attacks, particularly data hijacking or alteration**. Such attacks could pose significant risks to the system's accuracy and reliability, potentially leading to adverse consequences for air traffic controllers (ATCs) if they blindly follow incorrect suggestions.

Data hijacking refers to unauthorised access and interception of data transmitted between the intelligent assistant system and ATCs. Attackers with malicious intent could intercept and manipulate the data exchanged during communication, leading to the presentation of erroneous aircraft sequences to the ATCs. This manipulation may be subtle and difficult to detect, causing ATCs to unknowingly implement flawed instructions.

Data alteration involves unauthorised modification of data stored within the intelligent assistant system. Attackers could potentially modify critical information related to runway utilisation or aircraft sequencing algorithms, resulting in the generation of misleading suggestions. This alteration may not be immediately apparent, leading ATCs to rely on inaccurate recommendations and potentially causing disruptions to air traffic operations.

If ATCs blindly follow incorrect suggestions due to data hijacking or alteration, air traffic control operations may encounter disruptions, delays, or even safety risks. A series of misinformed decisions could lead to unintended aircraft conflicts, inefficient use of runways, and increased workload for ATCs.

In addition to these direct consequences, the potential adversarial, critical, or damaging effects of outages, attacks, misuse, or threats associated with the intelligent assistant system must also be considered. A successful cyberattack on the system could lead to prolonged service outages, hindering

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

ATCs' ability to access vital information and support tools. This disruption might cascade into broader operational challenges for airports and aviation authorities, impacting flight schedules and passenger safety. The level of exposure of the intelligent assistant system to cyber-attacks is not yet fully determined. However, as a precautionary measure, the system is expected to operate in an offline mode, only connected to the simulator to obtain necessary data. This offline operation could contribute to enhancing security.

The security analysis also considers the impact of the intelligent assistant system on the rights of privacy, physical, mental, and moral integrity, as well as data protection. Air traffic controllers might have concerns regarding the system's potential to reveal their efficiency in performing their tasks.

Strong authentication mechanisms need to be in place to ensure that only authorised users, such as ATCs, can access and interact with the intelligent assistant system. This helps prevent unauthorised access and misuse of critical information. Conducting regular security audits and vulnerability assessments is vital to identify and address potential weaknesses in the intelligent assistant system. This proactive approach helps to stay ahead of potential security risks. Providing training to ATCs on recognizing and responding to potential security threats is essential. Making them aware of best practices for securely interacting with the intelligent assistant system empowers them to maintain a high level of security in their operations.

As the system's development is in its early stages, a comprehensive security analysis is still underway. However, the implementation of offline operation and the careful consideration of security controls are expected to contribute significantly to the overall security and integrity of the Intelligent Sequence Assistant throughout its deployment and operation in air traffic control tower operations.

6.5. Liability assessment

6.5.1. Legal considerations about the UC4

From the legal perspective, the UC4 is grounded on the liability regime provided for ACTOs, as primary users of the IA.

To provide air traffic control service, any ATCO unit has to be provided with information on the intended movement of each aircraft, or variations therefrom, and with current information on the actual progress of each aircraft. In accordance with this information, the operators have to determine the relative positions of known aircraft to each other and issue clearances and information for the purpose of preventing collisions between aircraft under its control. In doing this, information on aircraft movements, together with a record of air traffic control clearances issues to such aircraft, shall be so displayed as to permit ready analysis in order to maintain an efficient flow of air traffic with adequate separation between aircrafts.

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

6.5.2. Actor-based liability analysis

Considering the role and tasks of the ATCO, the liability analysis is based on a specific legal regime. Some of these considerations were already anticipated in the analysis of the UC3.

In this regard, the civil liability regime should be related to the contractual relationship between employer and employee, coupled with the professional insurance required by law. The criminal liability outline, instead, would be deeply impacted by the accountability duties of this category of actors. Task responsibilities, therefore, should be considered beyond their nominal value and generally extended to the entire procedures, considered as a whole. This would include a general supervisory duty on the appropriate performance of other subjects' tasks (e.g., PIC, ground handlers, etc.).

In light of the above, as anticipated in the UC3 assessment, liability hypothesis for the ATCO can be confirmed if the following conditions are jointly satisfied:

- there is an **injury** to a legally protected interest;
- there is **careless behaviour** of the person at stake; and
- there is a **(causal) relation** between the behaviour and the injury.

The criteria previously described for assessing the lack of proper duty to care, however, are set out with greater detail by international and national laws (such as navigation codes) public or private standards and regulations, or even customs. However, you have to bear in mind that ATCOs professional standards are carefully and systematically defined by the Air Traffic Control Procedures Manual. In fact, in several legal cases, the controller who has fulfilled his obligations (as imposed in the manual) has been cleared of further liability.

Beyond the ATCO's specific tasks responsibilities, in this case, s/he may also be considered accountable for the tasks performed by other operators and subject to a general duty of care and proactive attitude to monitor and prevent potential risk situations.

However, it is essential to underline that **the ATCO – as an end-user of the IA – needs to be enabled to properly perform her/his tasks with the support of the new tool**. Therefore, the actor-based liability hypothesis will be confirmed only if the causal link between the ATCO's conduct and the injury is fully attributable to this actor.

If the negative occurrence is also correlated to other factors not only correlated to the ATCO's behaviour, but there could also be a secondary liability hypothesis. Producers may be involved if the injury is due to product defects affecting the IA. If instead the problems were due to poor implementation plans (e.g., the quality of the products and procedures, as well as implementation and investment strategies), there might be an organisation liability hypothesis. Considering the evolutive nature and the possibility of customisation of some AI-enabled technologies over time, these different liability hypotheses for developers and producers and implementing organisations can coexist and have contributory nature.

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

6.5.3. Liability assessment results for the UC4

Generally, the liability analysis discloses a deep intrinsic correlation among producers, employer organisations and the ATCO strategies and behaviours.

In particular, **the organisations (e.g., ANSPs) are responsible for all the organisational aspects of these innovations.** They are indirectly responsible for the behaviour of their employees in their interactions with the new tools and they have to ensure appropriate training and adequate usage conditions. Moreover, they are responsible for all the organisational aspects of these innovations, choosing only those products or solutions adequate for their operative purposes and tasks and reviewing all the internal policies and procedures impacted by the innovation.

The introduction of the ISA may raise some specific liability risks. From a material perspective, this tool should support the performance of relevant ATCO' tasks. According to the current description of UC4, this IA is primarily intended to support the arrangement of landing and take-of sequence. However, this would not be the only task delegated to the assistant. Indeed, this would have a relevant role in the management of the runways and taxiway, contributing to the tower control intended as a whole.

The preliminary analysis showed how the implementation of the ISA would not drastically change the current routing of the ATCO in the performance of her/his tasks. As remarked by the UC owner, the ATCO is free to reject the sequences suggested by the IA. **Possible issues due to a poor collaboration between the IA and the human agent (and the consequent mistakes) would result in the clearance of a turn around and in some correlated delay, without raising more critical safety issues**.

In the transition phase, therefore, the introduction of this IA would not particularly increase the liability risks exposure of the ATCO. The main issue concerns the possible mistakes related to the use of a new tool, but these should be mitigated by appropriate training. These actors, indeed, will have to familiarise themselves with the use of a new tool, but they will be still able to manage the situation also without the support of this latter. In other words, relying on the current procedures and on their background, they would maintain an autonomous peer position in the interaction with the ISA.

However, **on a long-term perspective, the use of this tool claims the ATCO has a blind reliance on the suggestions provided by the IA. This situation may raise more relevant concerns.** Considering the liability regime applicable to the ATCO, there are issues related to the overconfidence and over-reliance on the support of the IA, and so an undue delegation of tasks for protective reasons. In case of accident, these scenarios may embed the risk of professional negligence, due to careless actions and/or careless omissions. **The ATCO would still remain the accountable actor for all the decisions made in the performance of her/his tasks. Therefore, looking at the present legal regime of this subject, the blind reliance in case of accident would not excuse the ATCO.**

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

6.6. Recommendations for the UC4

In light of the above, looking at the future development of the UC4, these are the main recommendations addressed to UC4 owners to mitigate the possible risks emerging from the preliminary assessment:

- **Fine-tuning the assistant to be more conservative in its recommendations so as to accommodate the ATCO's decision-making preferences and risk thresholds**.
- **To develop a HMI sensitive to ethics-based aspects related to the autonomy of the human agent as well as explainability needs**. More specifically, in UC4, it is advisable to pay particular attention to direct and indirect issues concerning human autonomy and dignity (free self-determination in decision-making) accountability (of decision-making and its consequences) fairness (accessibility and universal design, now and over time, also in light of the background of users) and societal well-being impact on work and skills. Where feasible, this care should take into account the operative needs related to situation awareness, better understanding of IA's decisions, user confidence in of deviating or overriding the proposed suggestions.
- **To ensure the ATCOs an adequate explanation, interpretation and comprehension of the results provided by the ISA**. These aspects should be considered in light of human reaction time and her/his operative time window. If feasible, the use of counterfactual evidence for rectifications could be useful. A clear definition of these factors can facilitate the allocation of liability between the actors involved.
- **To define adequate training for end users, ensuring they are well aware of the philosophy of the system as well as of its intrinsic limits**. The skilling process should include ethics-based aspects related to the autonomy of the human agent, and so grant the ATCOs will be able to manage the situation on her/his own, also following alternative procedures. The objectives and features of this training should be defined in parallel with the development of the IA, and progressively adapted according to its evolutions.

7. UC5 – IA to assist safety in data analysis in the airport

7.1. Concept description and possible scenarios

The UC5 considers the development of the **Airport Safety Watch (ASW), an IA able to leverage historical aviation data to enhance the safety of day-to-day airport operations**.

At the beginning of the HAIKU project, UC5 presented a unitary operative scenario where the ASW would have been primarily devoted to the predictive decision-making support for the airport safety team staff. However, over the last 12 months, that scenario has been split into two different ones: in the first case the ASW will analyse data providing meaningful insights for the improvement of safety of airport operations (i.e., incident reduction in the short-medium term); in the second, the IA will predict risky situations on a day-to-day basis, promptly suggesting the most adequate mitigations. Given the current level of definition and development of the UC5 scenarios, this first round of the

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

liability assessment focused on the first one, on data-analytics for short-medium term improvement of safety plans.

Against this background, the primary users of the tool are the team members of the airport safety staff, responsible for day-to-day safety management at the airport. More specifically, the UC takes into consideration the operative context of the London Luton Airport (LLA). Secondary users include other Safety Stack members, such as individual airlines, NATS, and Ground Handling Service Providers. However, for the sake of clarity and consistency, in this first release of the liability assessment, the attention converges on primary users only.

As outlined in D3.2, London Luton Airport is the duty-holder when it comes to safety, and as such it collects a vast amount of data from across the airport partners, creating over 50,000 entries to its safety management platform annually, all of which are categorised under their most relevant headings. The analysis of this data is undertaken manually. Right now, LLA cannot easily exploit all of this data, but with AI there is the potential to identify which of their efforts produce the best results. LLA would expect the insights from the application of AI, informed by expert human users, to lead to better approaches to incident reduction, as well as enhancing safety data collection, categorization, analysis and visualisation, so that they (and the entire Stack community) can better learn from the data and team up with the AI for more effective ways to reduce the frequencies of key incident classes.

In light of the above, **the ASW AI-based tool should be able to anticipate, react to, and mitigate emerging safety threats and hazards at the airport (airside). By continuously analysing operational data the IA will identify threats that are likely to occur in current operations. The IA will also enable deep dive analysis of the causes and contributory factors of incident types, enabling the identification of new solutions to reduce their risk**. The idea is to have an ASW to flag actionable safety intelligence to the airport operational community in real-time. ASW will notify the user every time there is something that requires attention and the prompt adoption of risk avoidance strategies.

In this regard, ASW would be active 24/7 for continuous observation (hence ‘safety watch’) as well as periodically to address particular incident occurrence types. The data sources include all weather, safety and operational data (principally traffic movements and aircraft and vehicle characteristics) at the airport as well as human performance data (e.g., length of time on shift). In principle, ASW could be operated by using a classic keyboard/display interaction for safety specialists. For other staff they will generally receive alerts via hand-held devices (smartphones and tablets where these are allowed).

ASW can be considered as an ‘oracle’ that can be consulted by safety staff. Interaction will mainly be in the form of directed queries from human staff to better understand safety alerts and results of deep dive analysis. The use of this tool should be always coupled with alternative tools and procedures, with time to analyse and compare the provided suggestions/alerts.

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

In light of the above and according to the Human-AI Teaming Types & Digital Assistants categories developed HAIKU WP3 (Table 1) the tool should act as an **informer**. Indeed, the interaction will mainly be in the form of directed queries from human staff to better understand safety alerts and results of deep dive analysis.

7.2. HF assessment

The HF Assessment for the ASW AI-based tool focuses on the collaboration between the Intelligent assistant and the airport safety staff at London Luton Airport (LLA). The Intelligent assistant is fixed once deployed and collaborates actively with its developers through continuous feedback to improve its performance. The primary goal of the collaboration is to predict and mitigate safety risks related to incorrect taxiing and selection pushback errors.

The collaboration between the human and the Intelligent assistant is repeated over time, and the interaction involves taking turns. The Intelligent assistant proactively anticipates problems, communicating warnings and meaningful insights based on data analysis to the airport staff. **The human has full agency in making the final decisions, while the Intelligent assistant contributes more by predicting situations and calculating risks.**

The location and context of the collaboration involve co-located physical interaction between the human and the Intelligent assistant at LLA. Other Stack partners such as airlines, ground handling service providers, and air traffic control, will receive safety alerts through the airport community app or other direct communication sources via LLA, based on outputs from the ASW.

The consequences of the Intelligent assistant failing to perform as designed are not significant, as it maintains the same risk level as before. However, the benefits of the Intelligent assistant performing as expected are significant, aiming to reduce incidents and improve the overall safety risk picture.

Assessments of the collaboration's outcome are mainly subjective, conducted by the Luton Airport safety stack, the governing body for safety at the airport. Both the human and the Intelligent assistant are considered trusting and trustworthy, as long as the Intelligent assistant functions correctly.

The mode of interaction between the human and the Intelligent assistant is via a screen-based interface. The system proactively spots possible problems and communicates with the operator. It is highly predictable, and the confidence level of communication with the human is currently not implemented but is being considered for the future.

The Intelligent assistant is not human-like, and there are no significant anthropomorphic tendencies. The primary users of the system are adults, and there are no specific cultural consistencies/norms mentioned for those collaborating with the Intelligent assistant.

One potential issue is the over-reliance of the human staff on the Intelligent assistant's predictions and warnings. This may lead to complacency and reduced vigilance in their decision-making process, causing them to blindly follow the AI's suggestions without thorough verification or consideration of alternative options. Another potential issue is the **lack of confidence in the Intelligent assistant's**

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

predictions. If the system fails to effectively communicate confidence levels to the human operators, they may doubt the accuracy and reliability of the AI's insights, leading to mistrust in the system and potential disregard of important safety alerts. Continuous interaction with the Intelligent assistant, along with other operational tasks, may lead to cognitive overload for the human operators. Managing workload efficiently is crucial to avoid errors and fatigue during collaboration.

To address the potential issues, several mitigations could be put in place. Comprehensive training should be provided to the airport safety staff, emphasising effective collaboration with the Intelligent assistant. This training should include understanding the AI's limitations, interpreting its predictions, and critically evaluating its outputs. It is essential to highlight the importance of maintaining human agency in the decision-making process. Implementing a clear and transparent communication system that conveys the confidence levels of the Intelligent assistant's predictions to the human operators can help build trust in the system. This transparency will enable the operators to make informed judgements about the AI's recommendations. Improving the human-machine interface to make it user-friendly and intuitive is crucial. Tailoring the interface to meet the specific needs of the airport safety staff and incorporating visual aids, clear indicators, and feedback mechanisms will enhance understanding and usability. Implementing workload management strategies, such as task prioritisation, workload sharing, and periodic breaks, will help prevent cognitive overload for the human operators during continuous interaction with the Intelligent assistant. Establishing a user feedback loop between the human operators and the Intelligent Assistant's developers is essential. Real-world experiences and user suggestions should continuously inform improvements to the AI's performance, ensuring its effectiveness and reliability.

Overall, the collaboration aims to enhance safety at LLA by providing prompt risk alerts and improving safety data analysis for better decision-making.

7.3. Safety assessment

Under normal operations, the specific risks, risk metrics, and risk levels of the intelligent assistant system in the use case were not explicitly defined. However, it was clarified that the system should have sufficient evidence to make confident predictions and that the **possibility of the system drawing attention away from another incident type was identified as a potential risk. Another potential issue includes over-reliance on the Intelligent Assistant's predictions,** leading to reduced human vigilance, and a lack of human verification, which may result in overlooking critical safety hazards.

In case of failures, potential issues include Intelligent Assistant malfunction, **generating incorrect safety predictions, and system downtime, leading to the loss of critical safety insights. It was stated that a low level of accuracy of the intelligent assistant system** could result in critical, adversarial, or damaging consequences. Furthermore, the importance of having the right data for the system to work properly was recognized. To address these, redundancy should be established in the AI system to ensure continuous operation, and regular data backups should be performed to minimise the impact of system downtime. Fail-safe mechanisms must be implemented to allow human operators to take

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

over in case of Intelligent assistant failures, ensuring uninterrupted safety monitoring and decision-making.

However, the assessment did not provide clear answers to questions related to risk mitigation strategies, safety critical levels of consequences, and specific contexts or conditions for ensuring accuracy and reliability during abnormal conditions.

7.4. Security assessment

The security assessment aimed to identify potential risks and vulnerabilities associated with the intelligent assistant system, particularly in the event of outages, attacks, misuse, or threats. Specific forms of potential attacks were not clearly defined. However, various threats must be considered. These attacks may include hacking attempts, data breaches, or denial-of-service attacks, with the intent of disrupting the ASW tool's functionality or manipulating its predictions. Unauthorised access is another concern, where individuals gaining unauthorised entry could misuse the AI system, leading to false safety alerts or unauthorised changes to safety protocols. Additionally, insider threats pose a risk, as malicious actions from internal staff with access to the ASW tool may compromise system integrity and sensitive data. Malicious actors might attempt to produce false or misleading safety predictions, jeopardising airport operations. To counter this, robust security protocols and access controls are crucial, alongside data validation techniques to ensure the AI receives accurate and reliable data.

The assessment did not provide detailed information about the potential adversarial, critical, or damaging effects of security breaches on the intelligent assistant system. It was mentioned that the system is not considered highly exposed to security threats, and the belief is that there is minimal interest in breaching it. The ASW AI-based tool may have software vulnerabilities or weak security configurations that could be exploited by attackers. Proper encryption and security measures must be in place to protect the storage and transmission of sensitive safety data, preventing unauthorised access. Additionally, human error can be a vulnerability, as improper handling of login credentials or accidental disclosure of sensitive information may lead to security breaches. Such breaches could disrupt airport operations, compromise safety protocols, or lead to the unauthorised release of sensitive safety data.

Regarding controls, the assessment did not explicitly outline the evaluation of the system's resilience against adversarial attacks or manipulation attempts. To enhance security and mitigate risks, several measures should be implemented. Robust access controls are crucial to limit system access to authorised personnel only. Multi-factor authentication adds an extra layer of security. Data encryption during storage and transmission ensures sensitive safety data remains protected. Regular updates and patching of the ASW tool and its underlying software help address known vulnerabilities. Security audits and penetration testing aid in identifying and addressing potential weaknesses. Measures to detect and prevent insider threats should be in place, such as monitoring user activities and access logs. Comprehensive security training for all staff using the ASW tool is essential to emphasise best

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

practices and the importance of safeguarding login credentials. Developing a robust incident response plan enables swift and effective action in case of a security breach. Compliance with relevant data protection and privacy regulations is vital to safeguard user data and maintain legal compliance.

7.5. Liability assessment

7.5.1. Legal considerations about the UC5

From the legal standpoint, the UC5 does not present particular issues for the users involved. The actors directly impacted by the introduction of SAW are subjects with a well-established legal regime, not particularly exposed to liability risks.

What can be more relevant in this case is the specific regulation provided by the ICAO and EC for airports safety and security. However, **the specific responsibilities and duties generally are on the shoulders of the managing organisations, as authors of the organisational decision to introduce a new AI-based tool for the performance of these data-driven safety assessments**.

The service providers have to develop and maintain a process to identify hazards associated with aviation products and services, with a combination of reactive and proactive methods (ICAO, Annex 19, Appendix 2, § 2.1). In the performance of their duties, to identify the accountable executive who, irrespective of other functions, is accountable on behalf of the organisation for the implementation and maintenance of an effective safety management system (SMS); clearly define lines of safety accountability throughout the organisation, including a direct accountability for safety on the part of senior management and identify the responsibilities of all the members of the management, as well as of employees, with respect to the safety performance of the organisation (ICAO, Annex 19, Appendix 2, § 1.2).

According to Annex 2 (1-8) these employees can be qualified as «safety-sensitive personnel», intended as «persons who might endanger aviation safety if they perform their duties and functions improperly, including, but not limited to, crew members, aircraft maintenance personnel and air traffic controllers».

7.5.2. Actor-based liability analysis

In light of the above, the actor-based liability analysis mainly focuses on the generic professional negligence standards applicable to managers and employees that have no specific accountability duties in aviation (i.e., PIC and ATCO). However, approaching the results, you have to bear in mind that the conditions for professional negligence have to be broadly and proactively interpreted, taking into account the no-specific accountability duties outlined by the ICAO Annex 19 and Annex 2 for safety services providers.

As already observed applying to the UAMC (UC3) a generic professional negligence regime, safety team staff has a generally flexible professional outline, with a non-specific accountability position and task responsibilities defined according to her/his specific duties. On these grounds, this actor should be

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

subject to a civil liability regime defined by her/his employment contract and covered by vicarious liability from her/his. From the criminal liability perspective, instead, s/he should be subject to general negligence standards and assessed according to a professional duty of care.

Figure 3 describes part of the map for the assessment of generic professional negligence in aviation, as applicable to the safety team staff.

A liability hypothesis can be confirmed if the following conditions are jointly satisfied:

- there is an **injury** to a legally protected interest;
- there is **careless behaviour**[2] of the person at stake; and
- there is a **(causal) relation** between the behaviour and the injury.

Some exceptions or counter arguments may be advanced, e.g., the fact that the person's behaviour lacked the will.

In light of the above, the safety team staff has no well-established accountability position. This is why it is reasonable to assume that, if qualified in these terms, **each operator is responsible for her/his own tasks but has context-limited proactive duties related to the procedures performed by actors not directly instructed by him/her**.

However, it is essential to underline that **the operator – as an end-user of the IA – needs to be enabled to properly perform her/his tasks with the support of the new tool**. Therefore, the actor-based liability hypothesis will be confirmed only if the causal link between the operator's conduct and the injury is fully attributable to this actor.

If the negative occurrences are also correlated to other factors not only correlated to the operator's behaviour, but there could also be a secondary liability hypothesis. Producers may be involved if the injury is due to product defects affecting the IA. If instead, the problems were due to poor implementation plans (e.g., the quality of the products and procedures, as well as implementation and investment strategies), there might be an organisation liability hypothesis. Considering the evolutive nature and the possibility of customisation of some AI-enabled technologies over time, these different liability hypotheses for developers, producers and implementing organisations can coexist and have a contributory nature.

[2] Careless behaviour may consist of a careless action or a careless omission. An individual's behaviour is careless when the person took action, and the action was careless. Carelessness is usually determined by assessing whether the action violates the standard of due care, which is the proper behaviour that a professional operator would have been required to follow in the given situation. Such expectations depend on the tasks assigned to the operators, as well as on international and national laws, public or private standards and regulations, or even customs. An individual's omission will be careless when the person failed to take action; the person had a duty to act; and the person's action would have prevented the injury. The content of the duty to act will depend on the tasks assigned to each member of the safety team, as well as on international and national laws, public or private standards and regulations, or even customs.

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

7.5.3. Liability assessment results for the UC5

The conditions outlined in the previous sections allow you to have a clearer view of the possible liability risks associated with the use of the ASW by safety team staff. As anticipated, the liability assessment has taken into particular consideration the use of the IA for data analytics for short- and medium-term improvement of safety plans.

At a glance, **safety team staff may experience possible liability risk exposure due to overconfidence and over-reliance on the ASW suggestions**. Careless actions and/or omissions, indeed, may be generated by a poor interpretation and/or understanding of the relative value of the information obtained and, as a consequence, by superficial amendments of the safety plans currently into force. However, the legal risks associated with the use of the tool need to be fairly contextualised. ASW will not be the only support tool for making these choices and the timeframe for decision-making allows an analytic and comparative understanding of the suggestions obtained by the tool. This is the reason that leads to believe that, **at this stage of the concept development, possible serious accidents (and the possibly resulting casualties, injuries and/or damages) will not be distinctively correlated to the use of the ASW per se**.

It is important to note that the liability analysis discloses a deep intrinsic correlation among producers, employer organisations and the staff members' strategies and behaviours. In particular, the organisations (e.g., airlines and air carriers) remain responsible for all the organisational aspects of these innovations. They are indirectly responsible for the behaviour of their employees in their interactions with the new tools and they have to ensure appropriate training and adequate usage conditions. Moreover, they are responsible for all the organisational aspects of these innovations, choosing only those products or solutions adequate for their operative purposes and tasks and reviewing all the internal policies and procedures impacted by the innovation.

7.6. Recommendations for UC5

In light of the above, looking at the future development of the UC5, these are the main recommendations addressed to UC5 owners to mitigate the possible risks emerging from the preliminary assessment:

- **To develop a HMI sensitive to ethics-based aspects related to the autonomy of the human agent as well as explainability needs**. More specifically, in UC5, it is advisable to pay particular attention to direct and indirect issues concerning human autonomy and dignity (free self-determination in decision-making) accountability (of decision-making and its consequences) fairness (accessibility and universal design, now and over time, also in light of the background of users) and societal well-being impact on work and skills. Where feasible, this care should take into account the operative needs related to situation awareness, better understanding of IA's decisions, and capability of overriding these decisions.

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

- **To ensure the human agents have an adequate explanation, interpretation and comprehension of the results provided by the ASW**. These aspects should be considered in light of human reaction time and her/his operative time window. If feasible, the use of counterfactual evidence (and how to raise counterfactual queries) for rectifications could be useful. A clear definition of these factors may benefit the liability allocation among the actors involved.
- **To define adequate training for end users, ensuring they are well aware of the philosophy of the system as well as of its intrinsic limits**. The objectives and features of this training should be defined in parallel with the development of the IA, and progressively adapted according to its evolutions.

8. UC6 – IA to monitor risk factor conditions associated with the indoor spread of infectious diseases in the airport

8.1. Concept description and possible scenarios

The UC6 aims to **develop COVAID, an IA that would tackle the critical issue of preventing the spread of airborne diseases in crowded areas, specifically airports. COVAID is built on a near real-time routing recommendation system powered by machine learning**. This IA aims to promote mobility as a service with the appropriate routing for the prevention of COVID-19 spreading, relying on statistics of person routing and air quality in the airport common areas.

The primary users of this tool should be the travellers who transit through the airport. However, over time the use of this IA may be extended, also including airport health and safety operators.

Processing data coming from the travellers' mobile phones and data coming from cameras and air quality sensors, the AI will recommend a routing scheme and the person will either accept or reject it based on her behavioural status. Should it not be accepted the route of the passenger will be then placed to the AI as it is.

This cutting-edge system accurately predicts congestion levels and enables passengers to avoid overcrowded areas by suggesting alternative routes that might be of interest to the passenger, thereby significantly reducing the risk of disease transmission while also improving the overall experience.

In this regard, the main tasks assigned to COVAID include:

- Routing of persons according to their own preferences and the preferences of other travellers.
- Classification of the likelihood of high or low chance COVID risk.
- Justification and recommendation of a sequence of places the travellers will visit in the airport based on an intelligent algorithm.

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

- Forecast of air quality in airport places for both the passengers and the operators.
- Statistics of person routing and air quality for airport health and safety operators

The human-AI team will be characterised by a recommendation system accessible via passengers' mobile phones and shop assistants' computer systems. The AI team will make recommendations, and the human team will assess and potentially intervene in the recommendations by selecting the visit route.

COVAID ensures enhanced safety, reduced disease transmission risk, and a seamless and stress-free travel experience.

In light of the above and according to the Human-AI Teaming Types & Digital Assistants categories developed HAIKU WP3 (Table 1) the tool should act as a secretary. Indeed, the AI will recommend a routing scheme and the person will either accept or reject it based on her behavioural status. Should it not be accepted the route of the passenger will be then placed to the AI as it is.

8.2. HF assessment

This HF analysis delves into the collaborative dynamics between humans and COVAID, an evolving Intelligent Assistant (IA) aimed at preventing the spread of airborne diseases in crowded areas in airports. Powered by machine learning, COVAID operates on a near real-time routing recommendation platform, continually evolving through model updates and user interactions, especially with passengers who provide valuable inputs and feedback.

The collaboration between developers and COVAID is characterised by complete interaction, facilitating ongoing refinement of capabilities and optimal performance. This continual collaboration allows for the incorporation of user preferences into COVAID's routing recommendations, ensuring that the IA aligns with user needs. As part of the validation phase, COVAID will be tested and utilised by individuals beyond the original developers, specifically passengers transiting through airports. This inclusive approach ensures a comprehensive evaluation of the IA's effectiveness and user experience from diverse perspectives.

The goals of the human-AI collaboration within COVAID are clear and focused, revolving around promoting mobility as a service and ensuring passenger safety by providing appropriate routing to prevent the spread of COVID-19.

This collaboration involves both motivational and intellectual aspects. COVAID not only directs passengers along potential routes but also raises awareness about the environment and enhances safety measures, aligning the goals of the IA and its users.

Empathy is not a prerequisite for the human-AI interaction to function as intended. Instead, the collaboration is primarily based on trust, where COVAID's recommendations are designed to be reliable and user-centric, fostering user engagement. COVAID's collaboration primarily involves passengers who interact directly with the IA through their mobile phones. There are no third-party

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

individuals or Intelligent assistant systems involved, and COVAID operates as a virtual entity within passengers' mobile devices, offering recommendations and guidance remotely.

Passengers are fully aware that they are interacting with COVAID, as they explicitly consent to the download and usage of the app on their mobile phones. This transparency ensures that users willingly engage with the IA.

The consequences of COVAID failing to perform as designed are considered low in the short term. However, in the long term, the potential spread of diseases within airports could have significant implications. Conversely, the benefits of COVAID performing as expected are considerable, including enhanced passenger safety, reduced disease transmission risk, and improved travel experiences. In addition to immediate benefits, the human-AI collaboration within COVAID has broader implications for the future of AI adoption in similar contexts, fostering increased user acceptance and trust in AI-powered systems.

The assessment of this collaboration involves multiple stakeholders, including developers, users (passengers), and health authorities, encompassing both subjective and objective components to gauge COVAID's performance and impact effectively.

COVAID aims to establish bidirectional trust between passengers and the IA, a crucial factor in ensuring user engagement and adherence to the IA's recommendations. COVAID's interactivity primarily occurs through screen interfaces on passengers' mobile phones, where the IA provides routing recommendations and guidance. COVAID's proactive nature continuously anticipates the next steps of the interaction based on user inputs, recalculating and providing new routes in response to changing conditions and passenger feedback. The predictability of COVAID is considered moderate due to the specific context of airports and the diverse nature of passenger preferences and behaviours. COVAID's decision-making process and inputs are communicated to passengers through the screen interface of their mobile phones. COVAID exhibits a quite human-like communication style, which enhances user engagement and facilitates seamless interactions.

COVAID primary users are travellers, at large. For this reason, particular attention should be paid to vulnerable users and data subjects, with efforts to promote inclusivity by incorporating speech capabilities for visually impaired users. Passengers interacting with COVAID possess general knowledge of airport operations and norms, as well as technology interaction experience.

COVAID faces several challenges that require careful consideration for its effective and responsible use. One such challenge is the reliance on user consent for data availability. COVAID's success hinges on users willingly downloading and using the app. To ensure user engagement, it is essential to communicate the benefits of COVAID clearly. **Addressing privacy concerns and offering incentives or rewards can further encourage user consent, maximising the effectiveness of the IA**. Another potential issue lies in biased recommendations. COVAID's machine learning algorithms might unintentionally learn biases from historical data, leading to skewed recommendations for certain users or demographics. Employing fairness-aware AI techniques becomes crucial to identify and

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

mitigate biases, ensuring that the system provides fair and inclusive recommendations. Regular evaluations help maintain the integrity of the IA's suggestions. User trust and acceptance are critical factors determining COVAID's success. Some users may be hesitant to fully trust an AI system for essential tasks like navigating through airports. To build user trust, COVAID's interface should be designed to be transparent, explainable, and human-like in communication. Providing clear explanations for recommendations and showcasing how user feedback influences the system's behaviour fosters confidence in the IA. Inadequate user understanding can lead to misinterpretation of COVAID's recommendations. Ensuring users comprehend how to effectively use COVAID through in-app support and explanations enhances their understanding and ability to make informed decisions based on the IA's recommendations. **Accessibility for diverse users is a crucial aspect to ensure inclusivity. COVAID should be thoughtfully designed to cater to the needs of visually impaired or differently abled users**. Incorporating features such as speech capabilities, large text options, and intuitive navigation makes COVAID accessible to all users, regardless of their individual requirements. Lastly, addressing the potential over-reliance on technology is paramount. Encouraging users to perceive COVAID as an assistive tool rather than a replacement for personal judgement is essential. Emphasising the importance of remaining vigilant and aware of their surroundings while using COVAID prevents users from blindly relying on the IA and mitigates the risk of unintended consequences.

8.3. Safety assessment

The safety assessment of COVAID focuses on identifying risks and safety measures in the context of its specific use case – preventing the spread of airborne diseases in crowded areas, particularly airports. The analysis encompasses three key aspects: the initial design analysis under normal operations, the analysis considering abnormal conditions, and the evaluation in faulted conditions.

Under normal operations, one of the potential risks identified for COVAID is the possibility of **people not using the app correctly**, which could lead to overcrowded areas within the airport. If users do not follow COVAID's recommended routes or fail to adhere to the guidance provided by the IA, there is a risk of congestion and clustering in certain areas, increasing the likelihood of disease transmission. In response to this risk, the development team is proactively working on incorporating waiting factors into COVAID's routing recommendations. By introducing waiting factors, COVAID can dynamically manage the flow of passengers, ensuring a more even distribution of travellers across different areas within the airport. This approach can help alleviate congestion and prevent the formation of overcrowded places, reducing the risk of disease spread. The implementation of waiting factors aims to optimise the movement of passengers, ensuring a balanced distribution throughout the airport's common areas. By strategically staggering the timing of recommended routes or suggesting temporary pauses at certain points, COVAID can effectively regulate passenger movement and prevent bottlenecks.

To ensure data quality, measures are being developed to continuously assess the input data to the intelligent assistant system. Objective measures to assess the number of people in a place are being considered to enhance data quality assessment. However, the concrete implementation of these

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

measures is still in progress. Monitoring and documenting the accuracy of COVAID is an integral part of the safety assessment. The team is currently working on defining objective measures to evaluate the system's accuracy in recommending routes for passengers. However, specific steps for monitoring and documentation have not been put in place yet.

Similarly, measures to continuously assess the quality of COVAID's output have not been fully established yet. The team is actively working on developing mechanisms to ensure reliable and high-quality recommendations, but concrete implementation is still pending.

The team acknowledges that **COVAID could be vulnerable to misuse or inappropriate use by certain individuals**, leading to unintended consequences and safety risks. For instance, if users intentionally provide false information or deliberately ignore COVAID's recommendations, it could result in the creation of overcrowded places within the airport. Such overcrowding may compromise social distancing measures and increase the risk of disease transmission. To address this risk, the development team is proactively working on implementing measures to detect and prevent misuse. They are exploring techniques to validate user inputs and behaviour to distinguish genuine interactions from intentional misuse. Additionally, the team might consider implementing warnings or alerts to discourage improper use and encourage compliance with COVAID's recommendations.

The safety assessment recognizes the potential consequences of **COVAID's failures or malfunctions on human safety, particularly concerning the spread of infectious diseases**. If COVAID provides inaccurate or erroneous recommendations, passengers may inadvertently be directed to high-risk areas, leading to a higher likelihood of disease transmission. To mitigate this risk, safety critical levels of consequences are being defined, taking into account the severity and likelihood of potential failures. The severity of consequences will be linked to the infection spreading factor, meaning that the assessment will consider how likely it is for COVAID's errors to result in significant disease transmission. By understanding the potential impact of different failures, the team can prioritise efforts to prevent and address critical issues. The team is developing a mechanism to continuously assess the technical robustness and safety of the intelligent assistant system. This mechanism aims to detect any changes or updates to COVAID that may impact its performance and safety, triggering a thorough review and validation process. Moreover, tested failsafe fallback plans are being put in place to handle errors or faults that might occur within the intelligent assistant system. These fallback plans will act as safety measures to ensure COVAID can continue to operate effectively even in the presence of unexpected issues.

8.4. Security assessment

The security assessment of COVAID involves a comprehensive analysis of the primary and supporting assets, potential threats and vulnerabilities, and the identification of controls to safeguard the intelligent assistant system.

In the safety assessment of COVAID, the identification of assets plays a crucial role in understanding potential risks and vulnerabilities. The primary assets identified are the users' mobile phones, which

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

serve as the interface for interacting with COVAID, and the server that hosts the IA system. These assets are central to the functioning of COVAID, facilitating the exchange of data and recommendations between the IA and users.

The secondary asset identified is the data collected and processed by the system. This includes user information, such as location data, preferences, and behaviour patterns, which are utilised by COVAID to provide personalised routing recommendations. While this data is valuable for enhancing the IA's performance, it also presents a potential risk if compromised.

The safety assessment recognizes various threats to these assets, including outages, cyberattacks, misuse, and potential threats associated with the intelligent assistant. Outages or disruptions in the server hosting COVAID could impact the availability and functionality of the IA, hindering its ability to deliver real-time routing recommendations to users.

Two specific forms of cyberattacks mentioned are Distributed Denial of Service (DDoS) attacks and breaching of mobile phones. DDoS attacks involve overwhelming the server with a massive influx of traffic, causing it to become inaccessible to legitimate users. Breaching of mobile phones refers to unauthorised access to users' devices, potentially compromising personal data and the integrity of COVAID's interactions with users. Both these types of attacks pose significant risks to the availability, integrity, and confidentiality of COVAID's assets, potentially leading to disruptions in service and unauthorised access to user data.

Regarding the potential adversarial, critical, or damaging effects, the safety assessment highlights the risk of COVAID providing incorrect routing recommendations, directing users to wrong places within the airport. Such incidents could lead to the spread of diseases, particularly in crowded areas, and have critical consequences for public health and safety. For instance, if COVAID inaccurately directs users to high-risk or overcrowded areas, it could lead to an increased likelihood of disease transmission among travellers. This could result in a significant outbreak or further exacerbate an existing public health crisis. To mitigate these risks, the development team is actively working on **implementing robust security measures to protect assets, such as the server and user data, from cyber threats**. This includes measures to prevent DDoS attacks and enhance the security of users' mobile devices through encryption and authentication protocols.

Considering the exposure to cyber-attacks, the security assessment acknowledges that the level of exposure depends on the number of people using the system. This factor makes it difficult to assess the system's exact vulnerability to cyber-attacks at this stage.

COVAID's impact on fundamental rights, including the right to privacy, physical integrity, mental integrity, and data protection, is a critical consideration in the security assessment. As COVAID deals with personal data to provide personalised routing recommendations, privacy and data protection become vital concerns that must be thoroughly addressed to ensure compliance with relevant regulations and protect users' rights. The security assessment recognizes the potential risks associated with handling personal data and the importance of safeguarding users' privacy and data integrity. It

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

emphasises the **need for robust measures to protect the right to privacy and prevent any unauthorised access, breaches, or leaks of sensitive information**.

In terms of identification of controls, the security assessment explores several measures aimed at enhancing the system's resilience against adversarial attacks or manipulation attempts. While penetration testing to evaluate the system's, resilience has not been performed yet, it is actively under consideration for future testing. Penetration testing can help identify vulnerabilities and weaknesses in the system's security, allowing for targeted improvements and strengthening the system's overall security posture. To ensure authorised access to COVAID's functionalities and data, the assessment proposes robust authentication and access control mechanisms. Token-based authentication is being considered for implementation, which can provide an additional layer of security by generating unique tokens for each user, ensuring only authorised users can access the system's features and data. Furthermore, the security assessment acknowledges the importance of detecting and mitigating privacy breaches and sensitive information leaks. Although mechanisms for detection and mitigation are still in the planning phase, the assessment recognizes their significance in protecting users' privacy and data integrity. These mechanisms will be further developed and incorporated into the system in later stages of development to address potential privacy-related risks effectively.

8.5. Liability assessment

8.5.1. Legal considerations about the UC6

The UC6 presents peculiar features compared to the others. Indeed, safety and health services in the airports are issues usually regulated on a national basis. Indeed, on these topics, the ICAO only provides general recommendations, as adopted by specific ICAO Public Health Corridors (PHC).

Generally, a PHC is formed when two or more States agree to mutually recognize the implemented public health mitigation measures on one or more routes between their States. To enable such mutual recognition, States are strongly encouraged to actively collaborate and share information with other States and enter bilateral or multilateral discussions with each other to implement PHCs in a harmonised manner and mitigate the spread of disease. These coordination and cooperation efforts and initiatives, of course, have had a primary relevance during the COVID-19 pandemics.

It is noteworthy how, among the 10 Principles for a Safe, Secure and Sustainable Recovery proposed by the ICAO and its Council Aviation Recovery Task Force (CART), the Organization explicitly recommended that «States and industry should use data driven systemic approaches to manage the operational safety-, security-, and health-related risks in the restart and recovery phases and adapt their measures accordingly» (principle 4).

It is reasonable to assume that these tools will be used also in future to address similar situations. However, **the contents of the related recommendations may vary according to the specific needs of each health crisis**. This is the reason why the analysis of COVAID takes into consideration this guidance, but then focuses on the main legal references applicable to this kind of IA. Indeed, there are general

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

rules that may define the legal regime of these tools in future in normal as well as crisis situations.

In light of the above, it is essential to remark on the role that the processing of personal data will have for the current functioning and use of COVAID. **Even if the data were processed in an anonymous format, the processing activities may have a significant impact on the data subjects** that, in principle, will be guided over the airport by the suggestions provided by the IA.

The main legal reference, therefore, is the EU General Data Protection Regulation (reg. (EU) 2016/679, also known as GDPR) which provides the general standards for the protection of personal data.

In this regard, according to the GDPR, personal data is intended as «any information relating to an identified or identifiable natural person ('data subject')» More specifically, «an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person» (reg. (EU) 2016/679, Article 4(1)). On the other hand, the definition of 'processing' includes «any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction».

Considering the UC6 scenario, this induces us to presume that travellers can be qualified as data subjects and, therefore, they are potentially entitled to exercise their rights for the protection of their personal data. On the other hand, the airport manager organisation (or its safety and health protection team) should be qualified as controller, bearing all the responsibilities and duties related to the protection of personal data.

From a mere compliance-based perspective, the development of COVAID poses some relevant issues, mainly related to the **correct implementation of the protection measures by design and by default**. However, the most critical aspect concerns the legal basis for the processing of the personal data necessary for the effective functioning of the IA.

Generally, the processing of personal data by an app available on users' smartphones is based on the consent of data subjects. This involved the voluntary download and use of the app by travellers. The basic condition for the function of COVAID is the quantity of data available, and so the effective use of the app by the travellers: the more they use it, the more effective the screening will be. However, if they are reluctant to this form of tracking, the functioning of the app would be implicitly undermined.

During the pandemic period, many apps were developed for contact tracing and tracking. However, the suggestion from the EU Data Protection Authorities generally discouraged the mandatory use of these tools; and this notwithstanding the critical situation. Notwithstanding the evident reasons for public health, the Supervisory Authorities remarked on the risks associated with the use of these tools

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

and thus warmed the free involvement of people in these projects, fostering public confidence in the use of data more than the compulsory participation for the protection of the general interests.

Beyond the design issues, indeed, these limitations could have a relevant impact on the liability regime of the airport managing organisation, mainly associated with the acceptable risks related to the processing of this data.

8.5.2. Actor-based liability analysis

In light of the above, the actor-based analysis only partially relies on the argumentation maps. Indeed, considering the users involved, travellers are not exposed to specific liability risks. On the other hand, the airport managing entity is basically responsible – as a controller – for the protection of personal data, according to enterprise liability.

Considering the introduction of COVAID, a primary liability hypothesis for the enterprise can be confirmed if the following conditions are jointly satisfied:

- there is an **injury** to a legally protected interest;
- there is a **causal link** between the activities or processes of the enterprise and the injury;
- and the operational activities or processes are inadequate to prevent that injury (‘**organisational or systemic fault**’).

Analogous considerations are valid for a secondary (vicarious) liability hypothesis related to the behaviour of the employee of the considered enterprise. Indeed, the enterprise is liable for vicarious liability if the following conditions are jointly satisfied: i.e., the employee is personally liable for negligence, and s/he acted within the scope of the employment.

8.5.3. Liability assessment results for the UC6

The above considerations suggest paying particular attention to all the technical and organisational choices underlying the development of COVAID, since these may have severe consequences on the liability regime of the controller. The introduction of COVAID would expose the airport managing organisation to the liability risks associated with any violation of data protection law. And these liability risks can be declined in civil, criminal and/or administrative liability, according to the legal regime applicable to the facts considered.

8.6. Recommendations for the UC6

In light of the above, looking at the future development of the UC6, these are the main recommendations addressed to UC6 owners to mitigate the possible risks emerging form the preliminary assessment:

- **To promote a transparent mapping of the actors involved in the use of COVAID**, as well as of the categories of data collected and processed and the purposes of each processing activity. These preliminary operations shall be compliant with the GDPR.

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

- **To develop a tool and HMI sensitive to ethics-based aspects related to the autonomy of the human agent as well as privacy and data protection and transparency explainability needs**. This approach should include appropriate measures of privacy by design and by default, as per the GDPR.
- **To introduce a set of implementation policy enabling**: (1) a necessity and proportionality test for the processing of data in the given context, according to specificities of the current scenarios; (2) the free use of the COVAID, and the provision of a free and unconditioned consent to the processing of personal data, ensuring the easy exercise of the right to withdraw consent after the ad hoc use of the app.

9. Final considerations

9.1. IAs characterisation in light of the SHS-L assessments

A comparative analysis of the maturity level of the UCs highlights differences amongst the concepts' definitions. In some cases, the UC owners already have a very clear understanding of the expectations about their concept, as well as on the nature and scope of the tasks delegated to the IA, in other situations the outlining of this big picture is still in progress. The discrepancies may have an impact on the characterisation process suggested by the European Aviation Safety Agency, on the definition of the automation levels, as well as on the proportionality and modulation of the applicable AI guidance.

The application of the EASA AI levels, as well as the future descriptions of the HAIKU UCs, may already benefit from the results obtained by these first validation activities. As the Agency specified (EASA, 2023, p. 24), this preliminary characterisation should be based on the notion of authority, intended as «the ability to make decisions and take actions without the need for approval from other agents». The correct use of this notion does not rely on the automation model but on the distribution of tasks between AI-systems and end-users in light of the HAT scheme. Therefore, **the results obtained by the application of the SHS-L validation framework to the UCs may provide relevant insights about the effective apportionment of authority between the IAs and the human end-users**.

Table 2 summarises these observations extrapolated from the SHS-L assessments, with a specific focus on the results provided by the HF analysis. The aim is to support the improvement of the concepts' definition and IAs characterisation, bolstering a in a systematic approach.

Table 2. IAs characterisation

	UC description of HAT	SHS-L observations	Suggested level
UC1	The AI assistant is an informer, a coordinator and an executor and is therefore proactively providing support to pilots.	The degree of agency is balanced with the ultimate decision-making power resting with the pilot. This lack of clarity on IA goals and objectives may affect pilots' trust in the	2B

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

	By collaborating with the pilots to make sense with the situation, the assistant could also be defined as a rescuer.	system and their willingness to rely on its assistance. Pilots need training to understand the AI system's limitations and potential errors, prevent overreliance on the IA and maintain an active role in critical decision-making processes. Confidence indicators for startle effect detection should be developed to enhance reliability and predictability and facilitate responsibility apportionment.	
UC2	The IA is an informer and a secretary and therefore provides information proactively and decision support on demand.	There is cognitive shared mental model empathy between humans and the AI, contributing to a proactive interaction pattern. The IA has limited agency, handling supervised tasks, while pilots retain full decision-making authority. Transparent feedback on the AI's recommendations or decisions is crucial to prevent uncertainty and over-reliance on the system	2A
UC3	The IA is expected to coordinate the operations in the city sky, as a coordinator (at higher levels of cognitive control). The IA is expected to automatically perform low-level control and communications as well as repetitive tasks, as an executor (at lower levels of cognitive control) The IA is expected to perform some observation and information tasks to establish compatible and shared situation awareness between the DUC and the UAMC, as an observer and informer.	HAT involves both intellectual and motivational aspects, combining concurrent collaboration with occasional turn-taking for specific situations or occurrences. Divergent interpretations of goals or priorities could lead to conflicting decisions. The UAM Coordinator might overly depend on the Intelligent assistant, leading to complacency or reduced situational awareness. Efforts are being made to avoid any human-like characteristics in the Intelligent assistant to prevent anthropomorphism and maintain a clear distinction between the roles of human and AI.	2B/3A
UC4	The ATC and the IA form a team that work together to safely maximise the use of the runway, especially during 'peak' hours. The IA will help the ATC make the best possible decisions. In light of this, the IA could be qualified as a coordinator.	HAT is designed for cooperation: the IA offers proactive suggestions; the final decision-making and action-taking remain within the purview of the human ATCOs.	2A

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

		ATCOs need training to avoid over-reliance, maintain situational awareness and independent decision- making. ATCOs must be prepared to take full control of the operations and manage air traffic safely without the AI's assistance. ATCOs should have the authority to prioritise their judgement if IA suggests a sequence conflicting with their observations or other systems	
UC5	It is a tool for the safety managers and safety analysts, allowing alerts to be transmitted to airside workers on the ground (and in departing aircraft). It can be considered as an 'oracle' that can be consulted by safety staff. Interaction will mainly be in the form of directed queries from human staff to better understand safety alerts and results of deep dive analysis. In light of this, the IA could be qualified as an informer.	The human has full agency in making the final decisions, while the Intelligent assistant contributes more by predicting situations and calculating risks. Over-reliance of the human staff on the Intelligent assistant's predictions and warnings may lead to complacency and reduced vigilance in their decision-making process, causing them to blindly follow the AI's suggestions without thorough verification or consideration of alternative options.	1B/2A
UC6	The AI will recommend a routing scheme and the person will either accept or reject it based on her behavioural status. Should it not be accepted the route of the passenger will be then placed to the AI as it is. In light of this, the IA could be qualified as a secretary.	The IA COVAID operates on a near real-time routing recommendation platform, continually evolving through model updates and user interactions, especially with passengers who provide valuable inputs and feedback. The predictability of IA is considered moderate due to the specific context of airports and the diverse nature of passenger preferences and behaviours.	2A

9.2. A comparative overview of the results of the SHS-L assessments

The results obtained by the assessments performed on each UC provide interesting insights for the future development of the HAIKU project. In particular, the recommendations suggested for the individual scenarios present certain common mitigations. However, the specificities of the different IAs considered in each CONOPS also needs specific complementary safeguards.

Generally, the UCs analysis outlines how in all the considered scenarios end users need an **adequate explanation**, interpretation and comprehension and **specific training**. These two suggestions are true both for professional end-users – e.g., PICs, ATCOs and airport staff – as well as for lay people that should use the IA on their personal smartphone while travelling. As EASA emphasised, the level of explanations (as well as the explainability of the adopted models) may vary in light of the specific operative needs of the target at issue. Nonetheless, these two recommendations document the

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

complementary nature of by design and by default mitigation measures, addressing the above-mentioned needs from over the whole technology life cycle. This means that, on the one hand, developers and producers have to take into the utmost consideration these aspects from the early stage of technology design, in order to avoid and prevent possible design and warning defects of the final product. On the other hand, user organisations have to provide thoroughly clear insights about the operational needs of their end-users, and elaborate an adequate training strategy, also in light of the intrinsic limitations of the solutions at stake. The contents of upskilling and reskilling programmes should be defined and agreed also by the final end-users, possibly defining the respective accountability duties and responsibilities.

Another intriguing aspect of the results obtained concerns the link between privacy and data governance issues as outlined in the recommendations provided for the UC1 and UC6, where the IAs also have to process personal data (also intended in the broad sense of the term) of the end users, these individuals may be subject for direct and/or indirect forms of subjection before the technologies they are using. This state of subjection may take many forms, also in terms of profiling, performance assessment and undue interference in decision making. In this regard, it is important to note that, if the use of this data is necessary for the smooth functioning of the IAs, developers, manufacturers and user organisations have to ensure that this information will not be used for third purposes. Indeed, this is an essential condition for trust building, especially in the workplace.

Eventually, there are recommendations regarding the **potential advantages of incorporating ethics-based suggestions for the future expansion of UCs**. It is vital to prioritise end-users who hold a position of accountability for decision-making depending on IA support as well as for collaboration among other actors. This recommendation proposes **taking a proactive and comprehensive approach to envisioning their future responsibilities in new processes, as well as the design of the interfaces intended for these parties** (EASA, 2023, pp. 41, 44-45). The Consortium recommends consulting the expected MOCs for AI-based systems (EASA, 2023, pp. 91-100) and following the ethics assessment process proposed by the EC and EASA. These issues will be addressed in more detail in the dedicated tasks within WP5 – Explainability in Human AI Teaming.

9.3. Methodological recommendations from the SHS-L assessment

The comparative overview of the results obtained from the SHS-L eventually provide some insights about the comprehensiveness and validity of the HAIKU validation framework.

As emphasised by the recommendations reported at the end of each UC assessment, **the use of five KPAs – i.e., safety, HF, security, legal compliance, and liability – facilitates a holistic approach to AI-based concepts analysis**. However, the updates occurred to the validation frameworks provided by EASA in its second concept paper questions the comprehensiveness of the initial five-layers model.

In particular, **the emphasis put on ethics-based assessment offered useful food for thought** (EASA, 2023, p. 40-45). This integration indeed adds something new, only partially considered by the considered KPAs. Bearing in mind the three pillars funding the EU AI trustworthiness (EC, 2020; HLEG-

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

AI, 2019) the traditional SHS assessments can be linked to robustness, legal compliance and liability to lawfulness. Ethics, instead, refers to something different, that invites to take a step forward when regulation and standards are uncertain, pursuing principles and requirements.

Especially when the use of the IA could involve higher responsibilities for the end-users involved, the suggestion is **to test the effectiveness of technical advance MOC in practice, assessing if the final results match with the general social expectations about the functioning and use of that tool**.

It is essential to note that, also in light of this finding, **ethics should not be intended as an "umbrella" KPA**. Instead, it should be intended **as an additional dimension to bridge the gap among the current AMCs and the objectives provided by the EU and EASA, making explicit the value-based that are not so clear when considering the other KPAs**.

In light of the above, the project validation framework should be enriched by this additional dimension, intended and addressed as explained. In this regard, the second release of the D7.2 is expected to include an ad hoc section for ethics-assessment methodologies, also drawing from the guidance meanwhile developed by EASA.

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

Annex A - References and bibliography

Bauranov, A., & Rakas, J. (2021). Designing airspace for urban air mobility: A review of concepts and approaches. *Progress in Aerospace Sciences*(105).

BFU. (2004, May 3). German Federal Bureau of Aircraft Accidents Investigation, Investigation Report, AX001-1-2/02, Uberlingen accident.

Cohen , A., & Shaheen, S. (2021). Urban Air Mobility: Opportunities and Obstacles. In *International Encyclopedia of Transportation* (p. 702-709).

Contissa, G., Laukte, M., Sartor, G., Schebesta, H., Masutti, A., Lanzi, P., . . . Tomasello, P. (2013). Liability and automation: Issues and challenges for socio-technical systems. *Journal of Aerospace Operations, 2*(1-2), 79-98.

EASA. (2020, February 7). Artificial Intelligence Roadmap 1.0. Human-centric appraoch to Ai in aviation. Cologne, Germany.

EASA. (2021, December 20). Concept paper: First usable guidance for Level 1 machine learning application. A deliverable of the EASA AI Roadmap. Cologne, Germany.

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

EASA. (2023, February). Concept paper: First usable guidance for Level 1 and 2 machine learning application. A deliverable of the EASA AI Roadmap. Cologne, Germany.

EC. (2022). Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive, or AILD) (COM/2022/496 final).

EC. (2022, September 22). proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive, shorter: AILD) (COM(2022) 496 final). Brussels.

EC, I. E. (2021, March 3). 'U-Space Regulatory Package', Presentation by the Commission and EASA.

Fiallos Pazmino, L. (2020). *The International Civil Operations of Unmanned Aircraft Systems under Air Law.* Wolters Kluwer.

Havel, B., & Sanchez, G. (2014). *The Principles and Practice of International Aviation Law.* Cambridge (UK): Cambridge University Press.

HLEG-AI. (2020). The Assessment List for a Trustworthy Artificial Intelligence. Brussels.

HLEG-AI. (2020). The Assessment List for a Trustworthy Artificial Intelligence. Brussels.

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

Hodgkinson, D., & Johnston, R. (2017). *International Air Carrier Liability.* New York (NY, USA): Routledge.

Horstmann, G. (2006). Latency and duration of the action interruption in surprise. . *Cognition & Emotion, 20(2),* , 242-273.

Huttunen, M. (2022). U-Space: European Union's Concept of UAS Traffic Management. In B. Scott, *The Law of Unmanned Aircraft Systems* (p. 97-112). Wolters Klower.

Koch, M. (1999). The neurobiology of startle. *Progress in neurobiology, 59(2),* 107-128.

Masutti, A. (2020). *Il diritto aeronautico.* Torino: Giappichelli.

Mendes de Leon, P., & Calleja Crespo, D. (2011). *Achieving the Single European Sky. Goals and Challenges.* Wolters Kluwer.

Milde, M. (2008). Liability for Damage Caused by Aircraft on the Surface – Past and Current Efforts to Unify the Law. *ZLW*, 532-557.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General . (s.d.).

Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union. (s.d.).

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act). (s.d.).

Schebesta, H. (2017). Risk Regulation Through Liability Allocation: Transnational Product Liability and the Role of Certification. *Air & Space Law, 42*(2), 107-136.

Scott, B. (2022). *The Law of Unmanned Aircraft Systems.* Wolters Kluwer.

Scott, B., & Trimarchi, A. (2020). *Fundamentals of International Aviation Law and Policy.* New York (NY, USA): Routledge.

Scott, B., Andritsos, K., & Trimarchi, A. (2022). What is in a Name: Defining Key Terms in Urban Air Mobility. *Journal Of Intelligent & Robotic Systems, 105*, 1-9.

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

Annex B - HAIKU liability framework

It is noteworthy how aviation law is primarily based on international treaties and conventions. These international law instruments, in particular, aim at fostering a regulatory approach as uniform as possible. The undersigned states have a prominent political and legal duty to transpose these shared norms and principles into their domestic legal system. In addition, national legislators and judges should promote a uniform and consistent application of these latter into their national legal practice.

For the purposes of HAIKU, the relevant references for the liability assessment are reported in the figure below (Figure B. 1).

Figure B. 1. International law documents applicable to HAIKU

Looking at the EU legal context, these considerations are further nuanced by the peculiar characteristics of this legal system. According to its founding principles, EU law aims at harmonising continental legislation and jurisprudence to strengthen the free movement of people, services, capital, and goods. Implementing these principles into the Single European Sky (SES) package, EU law often specifies international law principles, contextualising and detailing these later according to the objectives and purposes of the EU integration and political strategies. Nonetheless, legislation in this sector primarily focuses on private law aspects and uniform safety requirements. Liability issues usually remain within the competencies of Member States.

For the purposes of HAIKU, the relevant references for the liability assessment are reported in the figure below (Figure B. 2).

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

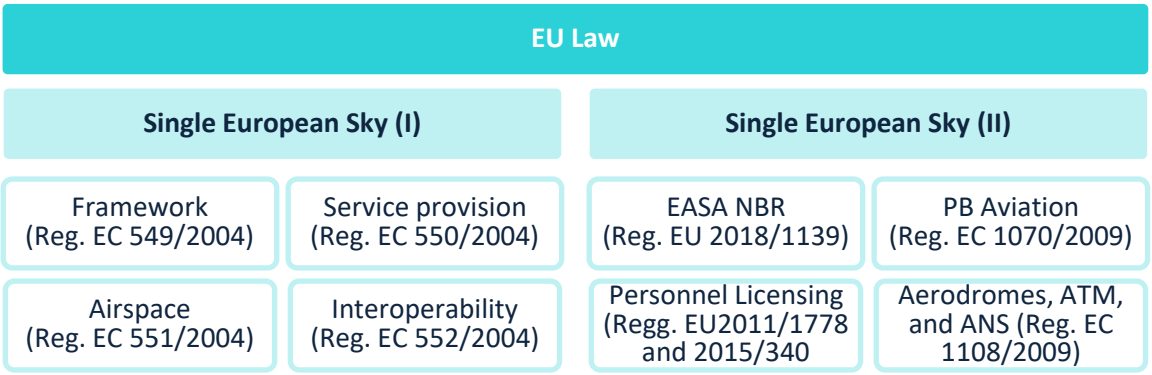

Figure B. 2. EU law applicable to HAIKU

The sources mapped in the tables above (Figure B. 1 and Figure B. 2) have general value. They apply to the aviation operations broadly intended, and implicitly outline the conditions that may envision potential liability risks exposures. A more detailed analysis of each of them is available in the Annex B. Moreover, a specific overview of each UC legal and regulatory framework and the liability regime of the actors involved will be provided in the following dedicated sections.

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

Annex C - UC1 SHS assessment grids

C.1 - Human Factors

Table C. 1. UC1 HF assessment grid

Category	Questions	Answers
I. Nature of Collaboration		
Stage of development or deployment	1. Is the Intelligent assistant fixed once deployed or evolving over time via model updates/continual interaction? 2. To what extent is there ongoing collaboration between the Intelligent assistant 's developer(s) and the system? [No collaboration, limited collaboration, moderate collaboration, active collaboration] 3. Is the Intelligent assistant system used by people other than the original developers?	1. Fixed. 2. Developers collect data that feed the model. 3. No
Goals	4. Are the goals of the human-AI collaboration clear or unclear? 5. What is the nature of the collaboration's goals? [Physical, knowledge/intellectual, emotional, and/or motivational in nature] 6. Is empathy a precondition for the human-AI interaction to function as intended? 7. Are the human and the Intelligent assistant system's goals aligned?	4. We are thinking of ways of interaction, but we still have to think about it. Maybe it will change in the future. I'm not sure. So maybe unclear. My position is we should let the human decide. So the AI may ask the pilot if he wants actions to be taken. But there should be a dialogue. But we still have to talk about this with pilots. 5. The purpose of the AI will be to help the pilot to unload emotions. It will also induce a kind of physical collaboration. For example, it can ask the pilot to make a deep breath. Therefore, there will be some actions from the pilot. 6. There is a kind of bio feedback. Given by the AI to the pilot about the pilot state so the purpose would be to help him to unload the emotions but also to have a clear count awareness of his state. 7. Should be
Interaction Pattern	8. Is the collaboration repeated over time or is it a one-time engagement? If over time, at what timescale? 9. Is the interaction concurrent – with both human and Intelligent assistant	8. it's a one-time engagement triggered by the detection of starter or surprise. We like to think that the assistant will help him to recover within the five minutes. 9. The essential part of the functioning is concurrent.

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

	contributing in parallel – or does it depend on taking turns?	
Degree of agency	10. Does the Intelligent assistant or human agent contribute more to the system's decision-making? Action-taking? 11. How much agency does the human have? The Intelligent assistant system? [None, limited, moderate, high, full]	10. We will use the term collaboration. Therefore, we don't want to say "more" The decision making, to be acceptable, should be led to the pilot in my own opinion. 11. The ultimate decision maker is the pilot
II. Nature of Situation		
Location and context	12. Are other people or other Intelligent assistant systems involved as third parties? 13. Are the human and Intelligent assistant agents co-located physically or virtually?	12. No 13. Physically
Awareness	14. Is the human likely aware that they are interacting with an Intelligent assistant system? 15. Does the human need to consent before interacting with the Intelligent assistant system?	14. Sure 15. No
Consequences	16. How significant are the consequences should the Intelligent assistant fail to perform as designed/expected? What are those consequences? [Low, moderate, high] 17. How significant are the benefits of the Intelligent assistant to the users should it perform as designed/expected? What are those benefits? [Low, moderate, high] 18. What are the potential consequences and benefits of the outcome of the collaboration? 19. What might be the broader impacts of the human-AI collaboration? 20. To what extent do typical users consider privacy and security when interacting with the Intelligent assistant agent? [Low, Moderate, High]	16. There should be no consequences. because we should give a free will of the pilot on the decision. 17. Well it's hard to answer because it completely depends on the situation. If it helps the pilot during a cruise phase where nothing happens for the next four hours the benefit might be low. On the other hand, if it helps the pilot during the final approach phase the benefit for the same thing might be significant. 18. A better situation awareness and fewer crashes. 19. A Cognitive assistant in the cockpit could provide support during phases where the pilot is not at 100%. this might lead to the introduction of AI collaboration in the cockpit. 20. We intend to use physiological data to feed the assistance so I think there might be some concern for physiological data.

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

Assessment	21. Who is the main party or individual assessing the nature and effectiveness of the human-AI collaboration? 22. Are assessments of the human-AI collaboration's outcome subjective or objective?	21. In the validation phase: psychologists, regulators. 22. Both; interviews and quantitative data.
Level of Trust	23. Are both the human and the Intelligent assistant trusting and trustworthy? AI trustworthiness can be defined broadly, driven by task competence, safety, authority, and authenticity, amongst other features (e.g., we know an AI comes from the same affiliation it claims to be from).	23. It will depend on the false positives, if the pilot trusts the system
III. AI System Characteristics		
Interactivity	24. What is the mode of the interaction between the two agents? [Via screen, voice, wearables, virtual reality, or something else] 25. Could the nature of the data that the Intelligent assistant system operates over impact its interactivity?	24. Screen 25. N/A
Adaptability	26. Is the Intelligent assistant system passively providing information or proactively anticipating the next steps of the interaction?	26. Well, the interaction with the intelligent assistant will be a kind of a step scenario where the pilot has to decide whether or not he continues. So, it's a kind of statement machine.
Performance	27. How predictable is the Intelligent assistant system? [Low, moderate, high] 28. Does the system often produce false positives? False negatives?	27. Predictable There should be no false positives. 28 There should be none.
Explainability	29. Can the Intelligent assistant system communicate its confidence levels to a human? 30. How does the Intelligent assistant system communicate its decision-making process and inputs to that decision-making process to the human?	29 At the moment we don't have any confidence indicators for the detection of the startle effect, but it could be interesting. 30. We were thinking of a classic Computer system, but a voice is possible. We were also thinking of giving an antropomorphic aspects to representing the pilot state to create a sort of interaction, but at the moment, we still have to decide it.

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

Personification	31. How human-like is the Intelligent assistant system? [Not very, moderately, or highly human-like] 32. How easily anthropomorphized is the Intelligent assistant system?	31. Maybe it will incorporate some aspects 32. Same as previous
IV. Human Characteristics		
Age	33. Is the person(s) collaborating with the Intelligent assistant system a child (under 18), an adult (18 - 65), or a senior (over 65)?	33. Adults
Differently-abled	34. Does the person collaborating with the Intelligent assistant have special needs or accommodations?	34. No
Culture	35. Are there cultural consistencies/norms for those collaborating with the Intelligent assistant system? 36. What level of previous technology interaction has the user(s) of the system had? [Low, moderate, high]	35. The general norms applied in the cockpit. Depending on the specific country the pilots might have some different norms and reluctance in accepting the system's suggestions. 36. High

C.2 - Safety

Table C. 2. UC1 Safety assessment grid

Category	Question	Answers
Initial Design analysis under normal operations.	1. Did you define risks, risk metrics, and risk levels of the intelligent assistant system in the specific use case? 2. Did you define clear risk mitigation strategies to address the identified safety risks? 3. Did you put in place measures to continuously assess the quality of the input data to the intelligent assistant system? 4. Did you put in place a series of steps to monitor and document the intelligent assistant system's accuracy? 5. Did you put in place measures to continuously assess the quality of the output(s) of the intelligent assistant system?	1. It depends on how well the system works and the level of false positives. But nevertheless, the ultimate decision is on the pilot. 2. The pilot ultimate decision is the strategy to mitigate the errors. 3. N/A 4. N/A 5. N/A

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

Initial Design analysis considering abnormal conditions.	1. Did you identify the risk of possible misuse or inappropriate use of the intelligent assistant system? If yes, did you identify the possible consequences? 2. Did you identify the potential impact of the intelligent assistant system's failures or malfunctions on human safety? 3. Did you define safety critical levels of the possible consequences of faults or misuse of the intelligent assistant system in terms of severity and likelihood? 4. Could a low level of accuracy of the intelligent assistant system result in critical, adversarial, or damaging consequences? 5. Could the intelligent assistant system cause critical, adversarial, or damaging consequences (e.g., pertaining to human safety) in case of low reliability and/or reproducibility? 6. Did you identify whether specific contexts or conditions need to be considered to ensure accuracy and reliability?	1. No 2. Given that the pilot will have the final decision no new risks are introduced. 3. we are thinking of an assistant that will help the pilot to unload emotions or help the pilot to reconstruct his situation awareness. Therefore, there is no decision is made by the assistant. 4. N/A 5. N/A 6. It depends on the amount of time available to the pilot to react and thus trust the assistant.
Initial Design analysis in faulted conditions.	1. Did you evaluate the robustness and reliability of the AI system under different operating conditions and potential failure scenarios? 2. Did you develop a mechanism to evaluate when the intelligent assistant system has been changed to merit a new review of its technical robustness and safety? 3. Did you put in place tested failsafe fallback plans to address intelligent assistant system errors of whatever origin and put governance procedures in place to trigger them? 4. Did you put in place a proper procedure for handling the cases where the intelligent assistant system yields result with a low confidence score?	1. N/A 2. N/A 3. N/A 4. N/A

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

C.3 - Security

Table C. 3. UC1 Security assessment grid

Category	Questions	Answers
Identification of Primary and Supporting Assets.	1. Did you identify the primary and secondary assets that could be affected in the event of outages, attacks, misuse, or threats associated with the intelligent assistant?	1. Data as primary assets. Secondary assets: supporting tech like Bluetooth or Wi-Fi.
Identification of Threats/Vulnerabilities.	2. Did you define potential forms of attacks to which the intelligent assistant system could be vulnerable? 3. Did you define the potential adversarial, critical or damaging effects in case of outages, attacks, misuse or threats associated with the intelligent assistant? 4. Did you define how exposed is the intelligent assistant system to cyber-attacks? 5. Did you consider the impact of the intelligent assistant system on the right to privacy, the right to physical, mental, and/or moral integrity, and the right to data protection?	2. Not yet 3. The final decision is on the pilot, so the damaging effect are almost none. 4. Unlikely that they will attack this system since there are other things more important in the cockpit. 5. N/A
Identification of Controls.	6. Did you evaluate the intelligent assistant system's resilience against adversarial attacks or manipulation attempts? 7. Did you consider robust authentication and access control mechanisms to ensure only authorised users can interact with the intelligent assistant system? 8. Did you identify mechanisms to detect and mitigate potential privacy breaches or leaks of sensitive information by the intelligent assistant system? 9. Did you identify monitoring mechanisms to detect and respond to security incidents or breaches involving the intelligent assistant system? 10. Did you identify measures to ensure the integrity, robustness, and overall security of the intelligent assistant system against potential attacks over its lifecycle?	6. N/A 7.N/A 8. N/A 9. N/A 10. N/A

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

Annex D - UC2 SHS assessments grids

D.1 - Human Factors

Table D. 1. UC2 HF assessment grid

Category	Questions	Answers
I. Nature of Collaboration		
Stage of development or deployment	1. Is the Intelligent assistant fixed once deployed or evolving over time via model updates/continual interaction? 2. To what extent is there ongoing collaboration between the Intelligent assistant 's developer(s) and the system? [No collaboration, limited collaboration, moderate collaboration, active collaboration] 3. Is the Intelligent assistant system used by people other than the original developers?	1. Fixed. The model cannot be modified over time. 2. Moderate collaboration 3. Pilots
Goals	4. Are the goals of the human-AI collaboration clear or unclear? 5. What is the nature of the collaboration's goals? [Physical, knowledge/intellectual, emotional, and/or motivational in nature] 6. Is empathy a precondition for the human-AI interaction to function as intended? 7. Are the human and the Intelligent assistant system's goals aligned?	4. Somewhat clear. In one month, they will be better defined. 5. Understanding, decision making and action taking. Physical and knowledge 6. Empathy in the sense of cognitive shared mental models yes, not in the emotional sense. 7. Yes
Interaction Pattern	8. Is the collaboration repeated over time or is it a one-time engagement? If over time, at what timescale? 9. Is the interaction concurrent – with both human and Intelligent assistant contributing in parallel – or does it depend on taking turns?	8. Repeated 9. Taking turns but in a very reactive way
Degree of agency	10. Does the Intelligent assistant or human agent contribute more to the system's decision-making? Action-taking? 11. How much agency does the human have? The Intelligent assistant system? [None, limited, moderate, high, full]	10. Equivalent. It might depend on the task. 11. Human full agency, AI limited (supervised) agency predefined on some tasks
II. Nature of Situation		
Location and context	12. Are other people or other Intelligent assistant systems involved as third parties?	12. ATC personnel, operational control centre 13. Co-located physically

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

	13. Are the human and Intelligent assistant agents co-located physically or virtually?	
Awareness	14. Is the human likely aware that they are interacting with an Intelligent assistant system? 15. Does the human need to consent before interacting with the Intelligent assistant system?	14. Yes 15. Yes
Consequences	16. How significant are the consequences should the Intelligent assistant fail to perform as designed/expected? What are those consequences? [Low, moderate, high] 17. How significant are the benefits of the Intelligent assistant to the users should it perform as designed/expected? What are those benefits? [Low, moderate, high] 18. What are the potential consequences and benefits of the outcome of the collaboration? 19. What might be the broader impacts of the human-AI collaboration? 20. To what extent do typical users consider privacy and security when interacting with the Intelligent assistant agent? [Low, Moderate, High]	16. The consequences are low from a safety point of view in the short run, but in 2050, when it will be necessary to adopt AI system to resolve complex situations, the consequences might be high. 17. High: reduction of stress of the pilot, reduction of work overload, improvement of the situation awareness and decision making. 18. From the operational point of view: reduction of costs, safety improvements 19. Reduction of complexity of mission management 20. Low
Assessment	21. Who is the main party or individual assessing the nature and effectiveness of the human-AI collaboration? 22. Are assessments of the human-AI collaboration's outcome subjective or objective?	21. Pilots and Airlines 22. Objective
Level of Trust	23. Are both the human and the Intelligent assistant trusting and trustworthy? AI trustworthiness can be defined broadly, driven by task competence, safety, authority, and authenticity, amongst other features (e.g., we know an AI comes from the same affiliation it claims to be from).	23. Yes. Trust is established with the interaction among the two.
III. AI System Characteristics		
Interactivity	24. What is the mode of the interaction between the two agents? [Via screen, voice, wearables, virtual reality, or something else] 25. Could the nature of the data that the Intelligent assistant system operates over impact its interactivity?	24. Screen 25. Yes

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

Adaptability	26. Is the Intelligent assistant system passively providing information or proactively anticipating the next steps of the interaction?	26. In the collaborative way it is proactive. In the cooperative is just reactive
Performance	27. How predictable is the Intelligent assistant system? [Low, moderate, high] 28. Does the system often produce false positives? False negatives?	27. Low, since if it was predictable the pilot would not need it 28. It should not
Explainability	29. Can the Intelligent assistant system communicate its confidence levels to a human? 30. How does the Intelligent assistant system communicate its decision-making process and inputs to that decision-making process to the human?	29. Not the case now 30. Through the user interface
Personification	31. How human-like is the Intelligent assistant system? [Not very, moderately, or highly human-like] 32. How easily anthropomorphized is the Intelligent assistant system?	31. Not very 32. Not at all
IV. Human Characteristics		
Age	33. Is the person(s) collaborating with the Intelligent assistant system a child (under 18), an adult (18 - 65), or a senior (over 65)?	33. Adult
Differently-abled	34. Does the person collaborating with the Intelligent assistant have special needs or accommodations?	34. The pilot needs some level of cognitive abstraction
Culture	35. Are there cultural consistencies/norms for those collaborating with the Intelligent assistant system? 36. What level of previous technology interaction has the user(s) of the system had? [Low, moderate, high]	35. There are some norms of operational concept and some norms related that the UC is based in Europe. 36. Low

D.2 - Safety

Table D. 2. UC2 Safety assessment grid

Category	Question	Answers
Initial Design analysis under normal operations.	1. Did you define risks, risk metrics, and risk levels of the intelligent assistant system in the specific use case?	1. It should not introduce further risks. Safety needs to assess what new risks are introduced with the new AI to the operations as they are now. 2. N/A

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

	2. Did you define clear risk mitigation strategies to address the identified safety risks? 3. Did you put in place measures to continuously assess the quality of the input data to the intelligent assistant system? 4. Did you put in place a series of steps to monitor and document the intelligent assistant system's accuracy? 5. Did you put in place measures to continuously assess the quality of the output(s) of the intelligent assistant system?	3. N/A 4. N/A 5. N/A
Initial Design analysis considering abnormal conditions.	1. Did you identify the risk of possible misuse or inappropriate use of the intelligent assistant system? If yes, did you identify the possible consequences? 2. Did you identify the potential impact of the intelligent assistant system's failures or malfunctions on human safety? 3. Did you define safety critical levels of the possible consequences of faults or misuse of the intelligent assistant system in terms of severity and likelihood? 4. Could a low level of accuracy of the intelligent assistant system result in critical, adversarial, or damaging consequences? 5. Could the intelligent assistant system cause critical, adversarial, or damaging consequences (e.g., pertaining to human safety) in case of low reliability and/or reproducibility? 6. Did you identify whether specific contexts or conditions need to be considered to ensure accuracy and reliability?	1. The pilot might be able to change some parameters or compromise the input data in a way to match his mental model but in doing so the system will not be able to fully help the pilot. 2. N/A 3. N/A 4. N/A 5. N/A 6. N/A
Initial Design analysis in faulted conditions.	1. Did you evaluate the robustness and reliability of the AI system under different operating conditions and potential failure scenarios? 2. Did you develop a mechanism to evaluate when the intelligent assistant system has been changed to merit a new	1. N/A 2. N/A 3. N/A 4. N/A

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

	review of its technical robustness and safety? 3. Did you put in place tested failsafe fallback plans to address intelligent assistant system errors of whatever origin and put governance procedures in place to trigger them? 4. Did you put in place a proper procedure for handling the cases where the intelligent assistant system yields result with a low confidence score?	

D.3 - Security

Table D. 3. UC2 Security assessment grid

Category	Questions	Answers
Identification of Primary and Supporting Assets.	1. Did you identify the primary and secondary assets that could be affected in the event of outages, attacks, misuse, or threats associated with the intelligent assistant?	1. N/A
Identification of Threats/Vulnerabilities.	2. Did you define potential forms of attacks to which the intelligent assistant system could be vulnerable? 3. Did you define the potential adversarial, critical or damaging effects in case of outages, attacks, misuse or threats associated with the intelligent assistant? 4. Did you define how exposed is the intelligent assistant system to cyber-attacks? 5. Did you consider the impact of the intelligent assistant system on the right to privacy, the right to physical, mental, and/or moral integrity, and the right to data protection?	2. N/A 3. N/A 4. N/A 5. N/A
Identification of Controls.	6. Did you evaluate the intelligent assistant system's resilience against adversarial attacks or manipulation attempts? 7. Did you consider robust authentication and access control mechanisms to ensure only	6. N/A 7. N/A 8. N/A

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

	authorised users can interact with the intelligent assistant system? 8. Did you identify mechanisms to detect and mitigate potential privacy breaches or leaks of sensitive information by the intelligent assistant system? 9. Did you identify monitoring mechanisms to detect and respond to security incidents or breaches involving the intelligent assistant system? 10. Did you identify measures to ensure the integrity, robustness, and overall security of the intelligent assistant system against potential attacks over its lifecycle?	9. N/A 10. N/A

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

Annex E - UC3 SHS assessment grids

E.1 - Human Factors

Table E. 1. UC3 HF assessment grid

Category	Questions	Answers
I. Nature of Collaboration		
Stage of development or deployment	1. Is the Intelligent assistant fixed once deployed or evolving over time via model updates/continual interaction? 2. To what extent is there ongoing collaboration between the Intelligent assistant 's developer(s) and the system? [No collaboration, limited collaboration, moderate collaboration, active collaboration] 3. Is the Intelligent assistant system used by people other than the original developers?	1. We believe now that it's evolving over time, but it's an open question. This is the hypothesis that we're working on, and then we'll see what the operators say about that when we test it. But that's the idea. 2. Moderate collaboration. It is crucial to have a specific role dedicated to system development. This role should encompass both online and offline responsibilities. 3. Currently it's not used by anyone else. It's only us at the moment. We haven't sold it or anything like that. But of course, it's not just going to be used by developers. It's going to be used by UAM Coordinators in the future if such a concept were to possibly be used.
Goals	4. Are the goals of the human-AI collaboration clear or unclear? 5. What is the nature of the collaboration's goals? [Physical, knowledge/intellectual, emotional, and/or motivational in nature] 6. Is empathy a precondition for the human-AI interaction to function as intended? 7. Are the human and the Intelligent assistant system's goals aligned?	4. I think the goal is clear. I mean, we know what the Intelligent assistant will assist humans with. In our operational scenario, I say clear. We think the goal is that the Intelligent assistant would reduce workload and also maintain the flow of the operations for Human, the UAM Coordinator, to maintain all the safety and actually manage everything. And if we are talking about the scenario in 2050, then we think that the automation, the AI would kind of do everything for the Human, and the Human is just the rollback.

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

		5. The essence of the collaboration lies in maintaining control. It revolves around shared control, that's all there is to it. So, I don't want to use any other words here because it will add ambiguity. 6. I say not really. Shared goal is sufficient, but that's a precondition. Alignment, as outlined in number seven, is crucial for successful collaboration. Without alignment, the collaboration is unlikely to be effective. So that would be the extent of empathy. At the very least, alignment is necessary, but it's also crucial to strike a balance in the degree of trust among the collaborators. 7. Their goals have to be aligned.
Interaction Pattern	8. Is the collaboration repeated over time or is it a one-time engagement? If over time, at what timescale? 9. Is the interaction concurrent – with both human and Intelligent assistant contributing in parallel – or does it depend on taking turns?	8. Repeated. I think it's continuous collaboration. It doesn't end. I mean, as long as the UAM operations is up and running so the collaboration continues endlessly. 9. The nature of collaboration in this context is a mixed hybrid approach. It involves both concurrent collaborations, as well as turn-taking when specific collaborative discussions or actions are required. During these instances, collaborators come together to address particular occurrences or situations. However, once the collaborative task is resolved, the collaboration reverts to a concurrent mode, with the assistant providing support throughout the process.
Degree of agency	10. Does the Intelligent assistant or human agent contribute more to the system's decision-making? Action-taking? 11. How much agency does the human have? The Intelligent assistant system? [None, limited, moderate, high, full]	10. I guess more is a bad word here. Yeah, I don't think we can answer this question because it's not more or less, it's a question of what here. 11. I guess it is also something that we will evaluate and assess in the scenario. We do

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

		not know yet and we will not really assess it using the word agency. So, I guess not how much agency, but what kind of autonomy does it have and how does it change with time?
II. Nature of Situation		
Location and context	12. Are other people or other Intelligent assistant systems involved as third parties? 13. Are the human and Intelligent assistant agents co-located physically or virtually?	12. Third parties for sure, because there can be a pilot or operator of a whole drone fleet or something that the intelligent assistant can coordinate with. And whether they have digital assistance as well, who knows? We haven't gone into that complication at all. 13. It's a question we should discuss much later. We will not even talk about it in this project. I think it will have cybersecurity implications, I suppose, and many other complications, but it's not something we will work on.
Awareness	14. Is the human likely aware that they are interacting with an Intelligent assistant system? 15. Does the human need to consent before interacting with the Intelligent assistant system?	14. Yeah, for sure. Yes. The coordinator is definitely aware. 15. I guess this is a work system, right? So, they have to consent just like they consent to use any other work system. There will not be much of a difference there. They will certainly know that they're using this assistant or tool or whatever we're calling it, at least in our concept. I mean, you're delegating to something.
Consequences	16. How significant are the consequences should the Intelligent assistant fail to perform as designed/expected? What are those consequences? [Low, moderate, high] 17. How significant are the benefits of the Intelligent assistant to the users should it perform as designed/expected? What are those benefits? [Low, moderate, high] 18. What are the potential consequences and benefits of the outcome of the collaboration? 19. What might be the broader impacts of the human-AI collaboration?	16. So that depends on the scenario, right? and its impact depends on the specific context and the stage of the development cycle. When you rely on it more, it will matter more. So, this is hard to say. I suppose one would start with a low consequence setting. 17. One of the benefits of utilizing an assistant in this context is the ability to make decisions much faster than humans. This can result in more optimal decisions being made. The speed and optimality of the assistant's decision-making can be

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

	20. To what extent do typical users consider privacy and security when interacting with the Intelligent assistant agent? [Low, Moderate, High]	considered quite significant. To illustrate, let's take the example of an emergency reroute. With the assistance of the system, it becomes possible to keep other traffic flowing smoothly while quickly reaching the desired destination. The assistant can efficiently select the appropriate end spot and manage the process effectively. Without such assistance, the interruption in traffic would be much more extensive, requiring the closure of larger areas. Furthermore, depending on the level of automation and the sophistication of the system, it may even be impractical to have drone traffic at all without the assistant's support. 18. same as 17 19. More efficient decision making and traffic coordination. 20. It depends on a little bit who the users are. Right? I mean, the police might have different considerations from a pizza delivery or whatever. So, we don't know. It's an open question, but we think it can be variable, let's say that.
Assessment	21. Who is the main party or individual assessing the nature and effectiveness of the human-AI collaboration? 22. Are assessments of the human-AI collaboration's outcome subjective or objective?	21. I guess that all users will have some thoughts about it, right? I mean, even airspace users could have thoughts if they are always given a bad rerouting or something by the system. But the main party, I think, is the UIM coordinator. If we're going to select the person closest to it, I say the UIM coordinator. 22. Absolutely, in assessing the performance of the system, both objective measures and subjective feedback from people are important. Objectively, it is possible to measure the system's effectiveness through various Key Performance Indicators (KPIs) to gauge how well it is functioning and meeting the desired objectives. These KPIs can provide valuable insights into the

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

		system's performance and allow for quantitative assessment. However, it is equally crucial to consider the subjective aspect of people's experiences and perceptions. Gathering feedback from individuals regarding their preferences, satisfaction levels, and overall sentiment towards the system is valuable in understanding its impact and user acceptance. People's thoughts and feelings about the system play a significant role in determining its success and adoption. While the specific validation plan for the project is yet to be developed, in theory, a combination of objective KPIs and subjective feedback can provide a comprehensive evaluation of the system's performance, both in terms of measurable outcomes and user experiences.
Level of Trust	23. Are both the human and the Intelligent assistant trusting and trustworthy? AI trustworthiness can be defined broadly, driven by task competence, safety, authority, and authenticity, amongst other features (e.g., we know an AI comes from the same affiliation it claims to be from).	23. I'd say that's an assumption we have. So, we are not going to assume that any party has any malicious intent. We are assuming that they are both trying to solve the traffic situation according to the system goals. So, they are truly collaborating. So, there is no game, so to speak, involved.
III. AI System Characteristics		
Interactivity	24. What is the mode of the interaction between the two agents? [Via screen, voice, wearables, virtual reality, or something else] 25. Could the nature of the data that the Intelligent assistant system operates over impact its interactivity?	24. Screen 25. In the main scenario that is currently being focused on, the primary interaction is limited to the screen interface. However, it's possible to consider incorporating a radio call or voice interaction in specific cases or alternate scenarios where it may be relevant. This could involve voice-based communication for certain functionalities or tasks. So, while the primary emphasis is on the screen interface, there is a potential for incorporating voice-based interactions in specific contexts as needed.

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

Adaptability	26. Is the Intelligent assistant system passively providing information or proactively anticipating the next steps of the interaction?	26. Proactive, I guess it's not just passively presenting numbers. It presents plans and options, for example.
Performance	27. How predictable is the Intelligent assistant system? [Low, moderate, high] 28. Does the system often produce false positives? False negatives?	27. High. 28. We don't know yet.
Explainability	29. Can the Intelligent assistant system communicate its confidence levels to a human? 30. How does the Intelligent assistant system communicate its decision-making process and inputs to that decision-making process to the human?	29. It's not a question we're looking at the moment. Might come later in the project. It's an important question. 30. Indeed, the focus of the research is on a screen-based interface for presenting the ongoing process. The specifics of how this presentation will be accomplished are part of the research objectives. Determining what information will be presented, when it will be presented, and to whom it will be presented are crucial questions that need to be explored in this project. Understanding how to effectively utilize the screen-based modality and designing a presentation framework that optimally conveys the relevant information are key aspects that will be investigated as part of the research.
Personification	31. How human-like is the Intelligent assistant system? [Not very, moderately, or highly human-like] 32. How easily anthropomorphized is the Intelligent assistant system?	31. We would like to avoid any human actually actively like aspect. That's why we might not even call it an assistant. 32. Not at all.
IV. Human Characteristics		
Age	33. Is the person(s) collaborating with the Intelligent assistant system a child (under 18), an adult (18 - 65), or a senior (over 65)?	33. Adults.
Differently-abled	34. Does the person collaborating with the Intelligent assistant have special needs or accommodations?	34. There will probably be what we call a screening procedure. Preselected. Same as air traffic controllers, I'd say at the moment is what we assume. Same selection criteria.

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

Culture	35. Are there cultural consistencies/norms for those collaborating with the Intelligent assistant system? 36. What level of previous technology interaction has the user(s) of the system had? [Low, moderate, high]	35. We're assuming that it will be a high safety culture environment. 36. Highly trained.

E.2 - Safety

Table E. 2. UC3 Safety assessment grid

Category	Question	Answers
Initial Design analysis under normal operations.	1. Did you define risks, risk metrics, and risk levels of the intelligent assistant system in the specific use case? 2. Did you define clear risk mitigation strategies to address the identified safety risks? 3. Did you put in place measures to continuously assess the quality of the input data to the intelligent assistant system? 4. Did you put in place a series of steps to monitor and document the intelligent assistant system's accuracy? 5. Did you put in place measures to continuously assess the quality of the output(s) of the intelligent assistant system?	1. N/A 2. N/A 3. N/A 4. N/A 5. N/A
Initial Design analysis considering abnormal conditions.	1. Did you identify the risk of possible misuse or inappropriate use of the intelligent assistant system? If yes, did you identify the possible consequences? 2. Did you identify the potential impact of the intelligent assistant system's failures or malfunctions on human safety? 3. Did you define safety critical levels of the possible consequences of faults or misuse of the intelligent assistant system in terms of severity and likelihood? 4. Could a low level of accuracy of the intelligent assistant system	1. N/A 2. N/A 3. N/A 4. N/A 5. N/A 6. N/A

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

	result in critical, adversarial, or damaging consequences? 5. Could the intelligent assistant system cause critical, adversarial, or damaging consequences (e.g., pertaining to human safety) in case of low reliability and/or reproducibility? 6. Did you identify whether specific contexts or conditions need to be considered to ensure accuracy and reliability?	
Initial Design analysis in faulted conditions.	1. Did you evaluate the robustness and reliability of the AI system under different operating conditions and potential failure scenarios? 2. Did you develop a mechanism to evaluate when the intelligent assistant system has been changed to merit a new review of its technical robustness and safety? 3. Did you put in place tested failsafe fallback plans to address intelligent assistant system errors of whatever origin and put governance procedures in place to trigger them? 4. Did you put in place a proper procedure for handling the cases where the intelligent assistant system yields results with a low confidence score?	1. N/A 2. N/A 3. N/A 4. N/A

E.3 - Security

Table E. 3. UC3 Security assessment grid

Category	Questions	Answers
Identification of Primary and Supporting Assets.	1. Did you identify the primary and secondary assets that could be affected in the event of outages, attacks, misuse, or threats associated with the intelligent assistant?	1. N/A
Identification of Threats/Vulnerabilities.	2. Did you define potential forms of attacks to which the intelligent assistant system could be vulnerable? 3. Did you define the potential adversarial, critical or damaging effects in case of outages, attacks,	2. N/A 3. N/A 4. N/A 5. N/A

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

	misuse or threats associated with the intelligent assistant? 4. Did you define how exposed is the intelligent assistant system to cyber-attacks? 5. Did you consider the impact of the intelligent assistant system on the right to privacy, the right to physical, mental, and/or moral integrity, and the right to data protection?	
Identification of Controls.	6. Did you evaluate the intelligent assistant system's resilience against adversarial attacks or manipulation attempts? 7. Did you consider robust authentication and access control mechanisms to ensure only authorised users can interact with the intelligent assistant system? 8. Did you identify mechanisms to detect and mitigate potential privacy breaches or leaks of sensitive information by the intelligent assistant system? 9. Did you identify monitoring mechanisms to detect and respond to security incidents or breaches involving the intelligent assistant system? 10. Did you identify measures to ensure the integrity, robustness, and overall security of the intelligent assistant system against potential attacks over its lifecycle?	6. N/A 7. N/A 8. N/A 9. N/A 10. N/A

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

Annex F - UC4 SHS assessment grids

F.1 - Human Factors

Table F. 1. UC4 HF assessment grid

Category	Questions	Answers
I. Nature of Collaboration		
Stage of development or deployment	1. Is the Intelligent assistant fixed once deployed or evolving over time via model updates/continual interaction? 2. To what extent is there ongoing collaboration between the Intelligent assistant 's developer(s) and the system? [No collaboration, limited collaboration, moderate collaboration, active collaboration] 3. Is the Intelligent assistant system used by people other than the original developers?	1. Evolving over time 2. There will be further updates by the developers 3. Developers and air traffic controllers
Goals	4. Are the goals of the human-AI collaboration clear or unclear? 5. What is the nature of the collaboration's goals? [Physical, knowledge/intellectual, emotional, and/or motivational in nature] 6. Is empathy a precondition for the human-AI interaction to function as intended? 7. Are the human and the Intelligent assistant system's goals aligned?	4. Clear. The intelligent assistant will suggest sequence and it's up to air traffic control to follow it or not. In a similar way as a GPS works on your car. It provides you with the most efficient route where you can follow it or not. 5. Mostly intellectual. So, at the end, the assistant would provide with a sequence because it has bigger data. So, it's basically intellectual more than anything else. 6. The human needs to trust the system and there must be a shared understanding of the situation 7. Aligned
Interaction Pattern	8. Is the collaboration repeated over time or is it a one-time engagement? If over time, at what timescale? 9. Is the interaction concurrent – with both human and Intelligent assistant contributing in parallel – or does it depend on taking turns?	8. Repeated over time continuously. 9. They contribute in parallel
Degree of agency	10. Does the Intelligent assistant or human agent contribute more to the system's decision-making? Action-taking?	10. The decision making and the action taking fully fall on the human. The intelligent assistant only suggests.

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

	11. How much agency does the human have? The Intelligent assistant system? [None, limited, moderate, high, full]	11. The human does have full agency in the decision. It's up to the human always. Intelligent assistant here does not do anything rather than suggesting.
II. Nature of Situation		
Location and context	12. Are other people or other Intelligent assistant systems involved as third parties? 13. Are the human and Intelligent assistant agents co-located physically or virtually?	12. No 13. Co-located physically
Awareness	14. Is the human likely aware that they are interacting with an Intelligent assistant system? 15. Does the human need to consent before interacting with the Intelligent assistant system?	14. Yes, fully aware 15. No
Consequences	16. How significant are the consequences should the Intelligent assistant fail to perform as designed/expected? What are those consequences? [Low, moderate, high] 17. How significant are the benefits of the Intelligent assistant to the users should it perform as designed/expected? What are those benefits? [Low, moderate, high] 18. What are the potential consequences and benefits of the outcome of the collaboration? 19. What might be the broader impacts of the human-AI collaboration? 20. To what extent do typical users consider privacy and security when interacting with the Intelligent assistant agent? [Low, Moderate, High]	16. They could be moderate because if the human blindly follows the assistant' suggestions, and for whatever reason They are not accurate enough for, the ATC may find itself in a difficult spot. This can still happen now however, so no other risks are introduced. 17. Benefits are high. Reducing greatly the efforts of the ATC. 18. Reducing stress and improving decision making 19. There might be a blind reliance on the intelligent assistant with the possibility of de-skilling the ATC. 20. That's quite a hard question to answer. I would say moderate, because there might be some concerns about privacy and security in terms of How efficient are you? How good or bad are you performing compared to other colleagues? That's something that people might be a little bit concerned.
Assessment	21. Who is the main party or individual assessing the nature and effectiveness of the human-AI collaboration? 22. Are assessments of the human-AI collaboration's outcome subjective or objective?	21. ATCs 22. They will probably be a bit of both, but there will always be a subjective outcome. Because at the end, every controller is different. So yes, some may find it helpful some may find it less helpful. Even though

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

		the outcome of the assistance is objective because it's a number.
Level of Trust	23. Are both the human and the Intelligent assistant trusting and trustworthy? AI trustworthiness can be defined broadly, driven by task competence, safety, authority, and authenticity, amongst other features (e.g., we know an AI comes from the same affiliation it claims to be from).	23. They should be. That's an assumption.
III. AI System Characteristics		
Interactivity	24. What is the mode of the interaction between the two agents? [Via screen, voice, wearables, virtual reality, or something else] 25. Could the nature of the data that the Intelligent assistant system operates over impact its interactivity?	24. Screen 25. N/A,
Adaptability	26. Is the Intelligent assistant system passively providing information or proactively anticipating the next steps of the interaction?	26. Is proactively anticipating the next steps of the interaction., the assistant will always have a bigger picture of what's coming and it will recalculate proactively.
Performance	27. How predictable is the Intelligent assistant system? [Low, moderate, high] 28. Does the system often produce false positives? False negatives?	27. Predictable as long as the limits of the system are understood. 28. Can't tell now
Explainability	29. Can the Intelligent assistant system communicate its confidence levels to a human? 30. How does the Intelligent assistant system communicate its decision-making process and inputs to that decision-making process to the human?	29. It won't 30. Assistant just communicates the outcome, not the process
Personification	31. How human-like is the Intelligent assistant system? [Not very, moderately, or highly human-like] 32. How easily anthropomorphized is the Intelligent assistant system?	31. No 32. No
IV. Human Characteristics		
Age	33. Is the person(s) collaborating with the Intelligent assistant system a child (under	33. Adult

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

	18), an adult (18 - 65), or a senior (over 65)?	
Differently-abled	34. Does the person collaborating with the Intelligent assistant have special needs or accommodations?	34. No
Culture	35. Are there cultural consistencies/norms for those collaborating with the Intelligent assistant system? 36. What level of previous technology interaction has the user(s) of the system had? [Low, moderate, high]	35. No 36. The user is used to work in a similar system.

F.2 - Safety

Table F. 2. UC4 Safety assessment grid

Category	Question	Answers
Initial Design analysis under normal operations.	1. Did you define risks, risk metrics, and risk levels of the intelligent assistant system in the specific use case? 2. Did you define clear risk mitigation strategies to address the identified safety risks? 3. Did you put in place measures to continuously assess the quality of the input data to the intelligent assistant system? 4. Did you put in place a series of steps to monitor and document the intelligent assistant system's accuracy? 5. Did you put in place measures to continuously assess the quality of the output(s) of the intelligent assistant system?	1. A possible risk is related to the fact that the assistant might suggest a sequence that the ATC is not really comfortable in following since it might be too risky, but he/she might still follow the suggestion resulting in a difficult situation to handle. 2. The system's suggestion could be made adaptable to the ATC. if the ATC feels that the assistant is too aggressive, we could try to tune it to make it a little bit more conservative. 3. Actually, yeah, we will have a traffic control running different scenarios in the simulator. We will have Feedback from them. So, this is probably the best way to continuously assess the quality of the input. 4. To document the accuracy, there is a certain set of rules and things that this the system must follow. And that would end up having an outcome of a number of operations per hour. That's the best way to monitor whether the intelligent assistant is as accurate or not. 5. Same
Initial Design analysis considering abnormal conditions.	1. Did you identify the risk of possible misuse or inappropriate use of the intelligent assistant system? If yes, did you identify the possible consequences?	1. N/A 2. N/A 3. N/A 4. N/A

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

	2. Did you identify the potential impact of the intelligent assistant system's failures or malfunctions on human safety? 3. Did you define safety critical levels of the possible consequences of faults or misuse of the intelligent assistant system in terms of severity and likelihood? 4. Could a low level of accuracy of the intelligent assistant system result in critical, adversarial, or damaging consequences? 5. Could the intelligent assistant system cause critical, adversarial, or damaging consequences (e.g., pertaining to human safety) in case of low reliability and/or reproducibility? 6. Did you identify whether specific contexts or conditions need to be considered to ensure accuracy and reliability?	5. Worst case scenarios if there's an emergency and the assistant has not recognized that as such the system would probably keep suggesting a sequence that is no longer valid 6. N/A
Initial Design analysis in faulted conditions.	1. Did you evaluate the robustness and reliability of the AI system under different operating conditions and potential failure scenarios? 2. Did you develop a mechanism to evaluate when the intelligent assistant system has been changed to merit a new review of its technical robustness and safety? 3. Did you put in place tested failsafe fallback plans to address intelligent assistant system errors of whatever origin and put governance procedures in place to trigger them? 4. Did you put in place a proper procedure for handling the cases where the intelligent assistant system yields result with a low confidence score?	1. N/A 2. N/A 3. Fail sale failsafe fallback plans. No, I wouldn't say. there's, such a fallback plan because if the system fails the air traffic controller will remain the controller, controlling the same way they are doing now. So, since this is just an addition there is no need of fallback plans. 4. N/A

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

F.3 - Security

Table F. 3. UC4 Security assessment grid

Category	Questions	Answers
Identification of Primary and Supporting Assets.	1. Did you identify the primary and secondary assets that could be affected in the event of outages, attacks, misuse, or threats associated with the intelligent assistant?	1. Data as the primary asset
Identification of Threats/Vulnerabilities.	2. Did you define potential forms of attacks to which the intelligent assistant system could be vulnerable? 3. Did you define the potential adversarial, critical or damaging effects in case of outages, attacks, misuse or threats associated with the intelligent assistant? 4. Did you define how exposed is the intelligent assistant system to cyber-attacks? 5. Did you consider the impact of the intelligent assistant system on the right to privacy, the right to physical, mental, and/or moral integrity, and the right to data protection?	2. Hijacking or altering of data. 3. The system could suggest incorrect sequences creating problems for the ATCs if they follow blindly 4. N/A 5. ATCs might be concerned for the fact that the system might show how efficiently they work.
Identification of Controls.	6. Did you evaluate the intelligent assistant system's resilience against adversarial attacks or manipulation attempts? 7. Did you consider robust authentication and access control mechanisms to ensure only authorised users can interact with the intelligent assistant system? 8. Did you identify mechanisms to detect and mitigate potential privacy breaches or leaks of sensitive information by the intelligent assistant system? 9. Did you identify monitoring mechanisms to detect and respond to security incidents or breaches	6. We don't know yet because it's too early right now. I think the system would be completely offline always so that helps in terms of Security and it will only be connected to the simulator to get the data from it. Nothing more than that. 7. N/A 8. N/A 9. N/A 10. N/A

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

	involving the intelligent assistant system? 10. Did you identify measures to ensure the integrity, robustness, and overall security of the intelligent assistant system against potential attacks over its lifecycle?	

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

Annex G - UC5 SHS assessment grids

G.1 - Human Factors

Table G. 1. UC5 HF assessment grid

Category	Questions	Answers
I. Nature of Collaboration		
Stage of development or deployment	1. Is the Intelligent assistant fixed once deployed or evolving over time via model updates/continual interaction? 2. To what extent is there ongoing collaboration between the Intelligent assistant 's developer(s) and the system? [No collaboration, limited collaboration, moderate collaboration, active collaboration] 3. Is the Intelligent assistant system used by people other than the original developers?	1. Fixed. There is no real interaction or evolution over time. 2. Active collaboration. The collaboration is based on the continuous feedbacks of the system to the developers. 3. The LLA staff
Goals	4. Are the goals of the human-AI collaboration clear or unclear? 5. What is the nature of the collaboration's goals? [Physical, knowledge/intellectual, emotional, and/or motivational in nature] 6. Is empathy a precondition for the human-AI interaction to function as intended? 7. Are the human and the Intelligent assistant system's goals aligned?	4. Clear. They are clear because we're, we're focusing on incorrect taxing and selection pushback error. So, it's all about predicting those in a sufficient timeframe that the airport personnel can react to mitigate those risks. 5. It's warning, basically. It's warning that, you know, there's going to be an increased risk for one or more of these error types in the coming hours. So, I guess it's knowledge / Intellectual. 6. I mean if it really means that there is an overlap of the Mental Models. I think it's partly so 7. Aligned
Interaction Pattern	8. Is the collaboration repeated over time or is it a one-time engagement? If over time, at what timescale? 9. Is the interaction concurrent – with both human and Intelligent assistant contributing in parallel – or does it depend on taking turns?	8. Repeated. The system will run continuously and flag problems. 9. Taking turns. The system gives feedback the operator responds.
Degree of agency	10. Does the Intelligent assistant or human agent contribute more to the system's decision-making? Action-taking?	10. So the intelligent assistant is predicting the situation. And then the human agent decides whether to then release that alert

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

	11. How much agency does the human have? The Intelligent assistant system? [None, limited, moderate, high, full]	to the airport staff. So, the heavy lifting is as we say it's done by the intelligent assistant. In that sense, it's contributing more but there's a human in the loop who makes the decision. 11. Human full agency, intelligent assistant just flags problems. The human has full agency because they are in control, but then the intelligent assistant is as I say doing all the calculations which we think would be beyond the human.
II. Nature of Situation		
Location and context	12. Are other people or other Intelligent assistant systems involved as third parties? 13. Are the human and Intelligent assistant agents co-located physically or virtually?	12. So no other intelligent assistants on the Horizon. Third parties. Yes. I mean all the airlines, the ground handling services, the air traffic control. They would, they would receive the alert. 13. Co-located physically.
Awareness	14. Is the human likely aware that they are interacting with an Intelligent assistant system? 15. Does the human need to consent before interacting with the Intelligent assistant system?	14. Yes 15. No
Consequences	16. How significant are the consequences should the Intelligent assistant fail to perform as designed/expected? What are those consequences? [Low, moderate, high] 17. How significant are the benefits of the Intelligent assistant to the users should it perform as designed/expected? What are those benefits? [Low, moderate, high] 18. What are the potential consequences and benefits of the outcome of the collaboration? 19. What might be the broader impacts of the human-AI collaboration? 20. To what extent do typical users consider privacy and security when interacting with the Intelligent assistant agent? [Low, Moderate, High]	16. So the consequences are not significant because it's giving them a warning. So, if the systems fail to release a then the system is basically as it is today. So, there'll be no additional risk. It would simply be the same risk as we have today. 17. So the point of this system would be to reduce incidents to zero or at least reduce them significantly. So that would make them significant. 18. I mean that's an interesting one because you're getting many stakeholders to work together for safety. So, there is these additional benefits of the collaboration. 19. Same as 18 20. Low
Assessment	21. Who is the main party or individual assessing the nature and effectiveness of the human-AI collaboration?	21. The Luton Airport, safety stack. They're like the governing body for safety.

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

	22. Are assessments of the human-AI collaboration's outcome subjective or objective?	22. I think it's going to be initially subjective, but objective as well.
Level of Trust	23. Are both the human and the Intelligent assistant trusting and trustworthy? AI trustworthiness can be defined broadly, driven by task competence, safety, authority, and authenticity, amongst other features (e.g., we know an AI comes from the same affiliation it claims to be from).	23. So if the intelligent assistant is working, well then then there'll be no trust issues.
III. AI System Characteristics		
Interactivity	24. What is the mode of the interaction between the two agents? [Via screen, voice, wearables, virtual reality, or something else] 25. Could the nature of the data that the Intelligent assistant system operates over impact its interactivity?	24. Screen. We're talking about an alert which would tell you that in, in several hours, there's going to be an issue. 25. N/A
Adaptability	26. Is the Intelligent assistant system passively providing information or proactively anticipating the next steps of the interaction?	26. Proactively spots a possible problem and communicates to the operator.
Performance	27. How predictable is the Intelligent assistant system? [Low, moderate, high] 28. Does the system often produce false positives? False negatives?	27. High 28. We cannot tell at the moment
Explainability	29. Can the Intelligent assistant system communicate its confidence levels to a human? 30. How does the Intelligent assistant system communicate its decision-making process and inputs to that decision-making process to the human?	29. Not at this stage. We are thinking about it. 30. Message on a dashboard
Personification	31. How human-like is the Intelligent assistant system? [Not very, moderately, or highly human-like] 32. How easily anthropomorphized is the Intelligent assistant system?	31. No 32. No
IV. Human Characteristics		
Age	33. Is the person(s) collaborating with the Intelligent assistant system a child (under	33. Adult

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

	18), an adult (18 - 65), or a senior (over 65)?	
Differently-abled	34. Does the person collaborating with the Intelligent assistant have special needs or accommodations?	34. No
Culture	35. Are there cultural consistencies/norms for those collaborating with the Intelligent assistant system? 36. What level of previous technology interaction has the user(s) of the system had? [Low, moderate, high]	35. We don't think so

G.2 - Safety

Table G. 2. UC5 Safety assessment grid

Category	Question	Answers
Initial Design analysis under normal operations.	1. Did you define risks, risk metrics, and risk levels of the intelligent assistant system in the specific use case? 2. Did you define clear risk mitigation strategies to address the identified safety risks? 3. Did you put in place measures to continuously assess the quality of the input data to the intelligent assistant system? 4. Did you put in place a series of steps to monitor and document the intelligent assistant system's accuracy? 5. Did you put in place measures to continuously assess the quality of the output(s) of the intelligent assistant system?	1. I think if it fails, then it's doesn't increase risk since the operations would be as they are now. At the moment, the one risk we have identified: What if the system draws attention away from another incident type? It will draw attention on a problem in one area but there might be another problem happening in another area. 2. N/A 3. N/A 4. N/A N/A
Initial Design analysis considering abnormal conditions.	1. Did you identify the risk of possible misuse or inappropriate use of the intelligent assistant system? If yes, did you identify the possible consequences? 2. Did you identify the potential impact of the intelligent assistant	1. N/A 2. N/A 3. N/A 4. N/A 5. N/A 6. The right data. The system needs the right data to work properly.

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

	system's failures or malfunctions on human safety? 3. Did you define safety critical levels of the possible consequences of faults or misuse of the intelligent assistant system in terms of severity and likelihood? 4. Could a low level of accuracy of the intelligent assistant system result in critical, adversarial, or damaging consequences? 5. Could the intelligent assistant system cause critical, adversarial, or damaging consequences (e.g., pertaining to human safety) in case of low reliability and/or reproducibility? 6. Did you identify whether specific contexts or conditions need to be considered to ensure accuracy and reliability?	
Initial Design analysis in faulted conditions.	1. Did you evaluate the robustness and reliability of the AI system under different operating conditions and potential failure scenarios? 2. Did you develop a mechanism to evaluate when the intelligent assistant system has been changed to merit a new review of its technical robustness and safety? 3. Did you put in place tested failsafe fallback plans to address intelligent assistant system errors of whatever origin and put governance procedures in place to trigger them? 4. Did you put in place a proper procedure for handling the cases where the intelligent assistant system yields results with a low confidence score?	1. N/A 2. N/A 3. N/A 4. N/A

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

G.3 - Security

Table G. 3. UC5 Security assessment grid

Category	Questions	Answers
Identification of Primary and Supporting Assets.	1. Did you identify the primary and secondary assets that could be affected in the event of outages, attacks, misuse, or threats associated with the intelligent assistant?	1. N/A
Identification of Threats/Vulnerabilities.	2. Did you define potential forms of attacks to which the intelligent assistant system could be vulnerable? 3. Did you define the potential adversarial, critical or damaging effects in case of outages, attacks, misuse or threats associated with the intelligent assistant? 4. Did you define how exposed is the intelligent assistant system to cyber-attacks? 5. Did you consider the impact of the intelligent assistant system on the right to privacy, the right to physical, mental, and/or moral integrity, and the right to data protection?	2. N/A 3. N/A 4. We don't think the system is very exposed. We don't think people would be interested in breaching in the system. 5. N/A
Identification of Controls.	6. Did you evaluate the intelligent assistant system's resilience against adversarial attacks or manipulation attempts? 7. Did you consider robust authentication and access control mechanisms to ensure only authorised users can interact with the intelligent assistant system? 8. Did you identify mechanisms to detect and mitigate potential privacy breaches or leaks of sensitive information by the intelligent assistant system? 9. Did you identify monitoring mechanisms to detect and respond to security incidents or breaches involving the intelligent assistant system? 10. Did you identify measures to ensure the integrity, robustness, and overall security of the intelligent assistant system against potential attacks over its lifecycle?	6. N/A 7. N/A 8. N/A 9. N/A 10. N/A

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

Annex H - UC6 SHS assessment grids

H.1 - Human Factors

Table H. 1. UC6 HF assessment grid

Category	Questions	Answers
I. Nature of Collaboration		
Stage of development or deployment	1. Is the Intelligent assistant fixed once deployed or evolving over time via model updates/continual interaction? 2. To what extent is there ongoing collaboration between the Intelligent assistant 's developer(s) and the system? [No collaboration, limited collaboration, moderate collaboration, active collaboration] 3. Is the Intelligent assistant system used by people other than the original developers?	1. Evolving over time thanks to the inputs of the passengers 2. Complete interaction between the developers and the system 3. It will be tested in the validation by other people other than the developers. They will represent the end users.
Goals	4. Are the goals of the human-AI collaboration clear or unclear? 5. What is the nature of the collaboration's goals? [Physical, knowledge/intellectual, emotional, and/or motivational in nature] 6. Is empathy a precondition for the human-AI interaction to function as intended? 7. Are the human and the Intelligent assistant system's goals aligned?	4. The nature of the collaboration is quite clear. The system provides successful recommendation for the routing in the airport. 5. Motivational and intellectual as well since it directs people in possible paths, and it raises awareness so that the passenger understands better the environment. 6. There is bi-directional communication. The aim is to have a better collaboration because of the trust that will be wanted between them. So that the human will engage and will start trusting the system and the system will in turn give better suggestions because of the knowledge basis, when it is populated. 7. The goal of the system and that of the users is aligned.
Interaction Pattern	8. Is the collaboration repeated over time or is it a one-time engagement? If over time, at what timescale? 9. Is the interaction concurrent – with both human and Intelligent assistant contributing in parallel – or does it depend on taking turns?	8. Repeated over time where the users give continual feedback to the intelligent assistant. 9. Taking turns: it will resemble a chatbot.

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

Degree of agency	10. Does the Intelligent assistant or human agent contribute more to the system's decision-making? Action-taking? 11. How much agency does the human have? The Intelligent assistant system? [None, limited, moderate, high, full]	10. It's not a matter of more. In this case the users will have the ultimate say in the decision making and action taking. 11. The intelligent assistant can simply give suggestions; the users will have the agency to follow them.
II. Nature of Situation		
Location and context	12. Are other people or other Intelligent assistant systems involved as third parties? 13. Are the human and Intelligent assistant agents co-located physically or virtually?	12. No the users will be the main party. 13. The intelligent assistant will be present on the users 'phones
Awareness	14. Is the human likely aware that they are interacting with an Intelligent assistant system? 15. Does the human need to consent before interacting with the Intelligent assistant system?	14. Yes 15. They will consent to the download of the app
Consequences	16. How significant are the consequences should the Intelligent assistant fail to perform as designed/expected? What are those consequences? [Low, moderate, high] 17. How significant are the benefits of the Intelligent assistant to the users should it perform as designed/expected? What are those benefits? [Low, moderate, high] 18. What are the potential consequences and benefits of the outcome of the collaboration? 19. What might be the broader impacts of the human-AI collaboration? 20. To what extent do typical users consider privacy and security when interacting with the Intelligent assistant agent? [Low, Moderate, High]	16. We need to separate in short term and long-term consequences. In the short term the criticality of the app not working is low. In the long term, the consequences might be high because of the spread of a disease. 17. There are different significant benefits: users getting aware of the danger of overcrowded places, the reduction of the chances of being infected, minimising the risks of overcrowded areas and the reduction of waiting times. 18. Reducing the spread of possible diseases 19. Educational benefits: users getting acquainted with AI systems and collaborate with the system. 20. It might depend on age; young people might be less concerned than older users who might have more important information in their phones.
Assessment	21. Who is the main party or individual assessing the nature and effectiveness of the human-AI collaboration?	21. At first the developers who will validate the assistant, then the users and health authorities. 22. The metrics will be both objectives and subjective (users' feelings).

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

	22. Are assessments of the human-AI collaboration's outcome subjective or objective?	
Level of Trust	23. Are both the human and the Intelligent assistant trusting and trustworthy? AI trustworthiness can be defined broadly, driven by task competence, safety, authority, and authenticity, amongst other features (e.g., we know an AI comes from the same affiliation it claims to be from).	23. That is the aim. The bidirectional conversation will be based on trust since the system does not know the preferences of the users are true. On the other hand, the user will trust the system on the basis of the outcome.
III. AI System Characteristics		
Interactivity	24. What is the mode of the interaction between the two agents? [Via screen, voice, wearables, virtual reality, or something else] 25. Could the nature of the data that the Intelligent assistant system operates over impact its interactivity?	24. Screen 25. N/A
Adaptability	26. Is the Intelligent assistant system passively providing information or proactively anticipating the next steps of the interaction?	26. Proactive. The system will recalculate over time and suggest new routes based on the users' inputs.
Performance	27. How predictable is the Intelligent assistant system? [Low, moderate, high] 28. Does the system often produce false positives? False negatives?	27. Quite predictable given the nature of the context (airports) 28. N/A
Explainability	29. Can the Intelligent assistant system communicate its confidence levels to a human? 30. How does the Intelligent assistant system communicate its decision-making process and inputs to that decision-making process to the human?	29. We are thinking about this. The system should aim to do so. 30. Via screen
Personification	31. How human-like is the Intelligent assistant system? [Not very, moderately, or highly human-like] 32. How easily anthropomorphized is the Intelligent assistant system?	31. Quite human-like in terms of communication style. 32. We don't think it will be
IV. Human Characteristics		
Age	33. Is the person(s) collaborating with the Intelligent assistant system a child	33. 18-65 mainly

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

	(under 18), an adult (18 - 65), or a senior (over 65)?	
Differently-abled	34. Does the person collaborating with the Intelligent assistant have special needs or accommodations?	34. To make the system more inclusive it could have speech so that also blind people could use it.
Culture	35. Are there cultural consistencies/norms for those collaborating with the Intelligent assistant system? 36. What level of previous technology interaction has the user(s) of the system had? [Low, moderate, high]	35. General airport knowledge. 36. General phone usage knowledge.

H.2 - Safety

Table H. 2. UC6 Safety assessment grid

Category	Question	Answers
Initial Design analysis under normal operations.	1. Did you define risks, risk metrics, and risk levels of the intelligent assistant system in the specific use case? 2. Did you define clear risk mitigation strategies to address the identified safety risks? 3. Did you put in place measures to continuously assess the quality of the input data to the intelligent assistant system? 4. Did you put in place a series of steps to monitor and document the intelligent assistant system's accuracy? 5. Did you put in place measures to continuously assess the quality of the output(s) of the intelligent assistant system?	1. Risk of people not using the app correctly and thus creating overcrowded places. 2. We are working on it, trying to incorporate waiting factors to manage the dynamic flow. 3. Not yet 4. N/A 5. We are thinking of objective measures to assess the amount of people in a place.
Initial Design analysis considering abnormal conditions.	1. Did you identify the risk of possible misuse or inappropriate use of the intelligent assistant system? If yes, did you identify the possible consequences? 2. Did you identify the potential impact of the intelligent assistant system's failures or malfunctions on human safety? 3. Did you define safety critical levels of the possible consequences of faults or misuse of the intelligent assistant system in terms of severity and likelihood?	1. People not using it correctly might create overcrowded places. 2. The spreading of infectious diseases 3. The severity depends on the infection spreading factor. 4. Not yet 5. Increasing infection rates 6. N/A

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

	4. Could a low level of accuracy of the intelligent assistant system result in critical, adversarial, or damaging consequences? 5. Could the intelligent assistant system cause critical, adversarial, or damaging consequences (e.g., pertaining to human safety) in case of low reliability and/or reproducibility? 6. Did you identify whether specific contexts or conditions need to be considered to ensure accuracy and reliability?	
Initial Design analysis in faulted conditions.	1. Did you evaluate the robustness and reliability of the AI system under different operating conditions and potential failure scenarios? 2. Did you develop a mechanism to evaluate when the intelligent assistant system has been changed to merit a new review of its technical robustness and safety? 3. Did you put in place tested failsafe fallback plans to address intelligent assistant system errors of whatever origin and put governance procedures in place to trigger them? 4. Did you put in place a proper procedure for handling the cases where the intelligent assistant system yields result with a low confidence score?	1. It depends on the phones' connection and potential deadlocks in the software. 2. It depends on the results metrics of the recommendations. 3. N/A 4. N/A

H.3 - Security

Table H. 3. UC6 Security assessment grid

Category	Questions	Answers
Identification of Primary and Supporting Assets.	1. Did you identify the primary and secondary assets that could be affected in the event of outages, attacks, misuse, or threats associated with the intelligent assistant?	1. Primary assets: phones, server Secondary assets: data
Identification of Threats/Vulnerabilities.	2. Did you define potential forms of attacks to which the intelligent assistant system could be vulnerable? 3. Did you define the potential adversarial, critical or damaging	2. DDoS, breaching of phones 3. People getting sent to the wrong places and thus spreading a disease 4. It depends on the amount of people using the system, difficult to say now.

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

	effects in case of outages, attacks, misuse or threats associated with the intelligent assistant? 4. Did you define how exposed is the intelligent assistant system to cyber-attacks? 5. Did you consider the impact of the intelligent assistant system on the right to privacy, the right to physical, mental, and/or moral integrity, and the right to data protection?	5. It is dealing with personal data as well
Identification of Controls.	6. Did you evaluate the intelligent assistant system's resilience against adversarial attacks or manipulation attempts? 7. Did you consider robust authentication and access control mechanisms to ensure only authorised users can interact with the intelligent assistant system? 8. Did you identify mechanisms to detect and mitigate potential privacy breaches or leaks of sensitive information by the intelligent assistant system? 9. Did you identify monitoring mechanisms to detect and respond to security incidents or breaches involving the intelligent assistant system? 10. Did you identify measures to ensure the integrity, robustness, and overall security of the intelligent assistant system against potential attacks over its lifecycle?	6. No, but we are considering testing it with penetration testing. 7. We are thinking about token-based authentication 8. We will think about this in a later stage 9. N/A 10. N/A

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