

Applying Human-Centered Design: A Practical Case Study developing an AI-based Digital Assistant for Pilots

Amandine Mayima[1], Carmen Bejarano Espada[1], Anja Gampe[2], Anneke Hamann[2], Riccardo De Masellis[1], Andrew Fuchs[1], Melina Athanasiadou[1], Hector Palop[1] & Dave Watkins[3]

[1]Collins Aerospace, Applied Research & Technology, Cork, Ireland
[2]German Aerospace Center, Institute of Flight Guidance, Braunschweig, Germany
[3]Collins Aerospace, Applied Research & Technology, Utah, USA

Abstract

The CAIRDA (Collins Airborne Digital Assistant) is an AI-based cockpit prototype developed using a Human-Centered Design (HCD) process to support pilots in safety-critical environments. The iterative development process involved pilot workshops, simulator-based testing, and a final user study with pilots to ensure alignment with operational needs. Key design principles included explainability, usability, and workflow integration, with refinements made to address pilot feedback, balance automation with human oversight, and enhance trust. This work demonstrates the value of HCD in developing AI-enabled systems and offers a foundation for future advancements in cockpit technologies.

Keywords: Human-Centered Design, Commercial Air Transport Operations, Artificial Intelligence

1. Introduction

The aviation industry is increasingly exploring the use of artificial intelligence (AI) to enhance safety, efficiency, and operational performance across all facets of air traffic management (ATM) [1]. However, the integration of AI into the flight deck presents unique challenges, as pilots operate in a safety-critical environment where human–machine collaboration is paramount [2]. One promising approach to address these challenges is the development of AI-based Digital Assistants (DAs) that can support pilots by reducing workload, improving task efficiency, and enhancing decision-making processes [3], all while maintaining or improving current safety levels.

The Collins Airborne Digital Assistant (CAIRDA) is an AI-based cockpit-integrated prototype developed as part of JARVIS[1], a three-year EU co-funded project under the SESAR3 Joint Undertaking programme. The overarching goal of the CAIRDA is to leverage AI-based machine intelligence to support pilot decision-making, automate selected functions, and improve crew performance while ensuring that safety standards are maintained.

While the SESAR project explored the potential of various DAs to support pilot tasks, this paper highlights the outcomes and insights gained from applying a human-centered design process to the development of an AI-based Digital Assistant. The CAIRDA represents one of three assistant prototypes developed within the project, and this paper provides a detailed account of the iterative process, design principles, and outcomes associated with this specific assistant.

This paper describes the Human-Centered Design (HCD) process employed, described in [4], to develop the CAIRDA prototype and presents the findings of a user study conducted with 21 professional pilots in a laboratory environment. The study evaluated the prototype's impact on pilot workload, trust, usability, and situation awareness during simulated flight scenarios. The key contributions of this research include insights into the challenges and opportunities of integrating AI-based digital assistants into safety-critical aviation environments. Ultimately, this work lays the foundation for higher TRL developments and operational adoption of AI-based cockpit assistants in the future.

[1] https://research.dblue.it/jarvis/

2. Methodology

The development of the CAIRDA prototype followed an HCD process, ensuring that the system met the operational needs of pilots and the constraints of the aviation environment. The HCD method is a systematic, iterative design approach that places people—with their needs, abilities, contexts and behavioural patterns—at the heart of the development of products, systems or services. It is based on empirical user data and aims to optimise usability, functionality and user-friendliness through continuous evaluation and adaptation. The process typically follows the phases of analysis, design, evaluation and iteration, with user feedback integrated at every stage, to develop solutions that are not only technically feasible but also socially and contextually appropriate. From a scientific perspective, HCD is an evidence-based approach that utilises interdisciplinary methods (e.g. ethnography, usability testing, prototyping) to systematically account for human factors and thus improve the interaction between people and technology. This is in line with the principles of human factors research and participatory design. The process was iterative and consisted of three main activities: **Understanding the User and Environment**, **Visualizing and Producing Design Solutions**, and **Evaluating Designs and Iterating Solutions**. Additionally, explainability was embedded as a cross-cutting design principle to enhance trust and transparency in the system. Each step of the process is detailed in the subsections below and an overview is provided in Figure 1.

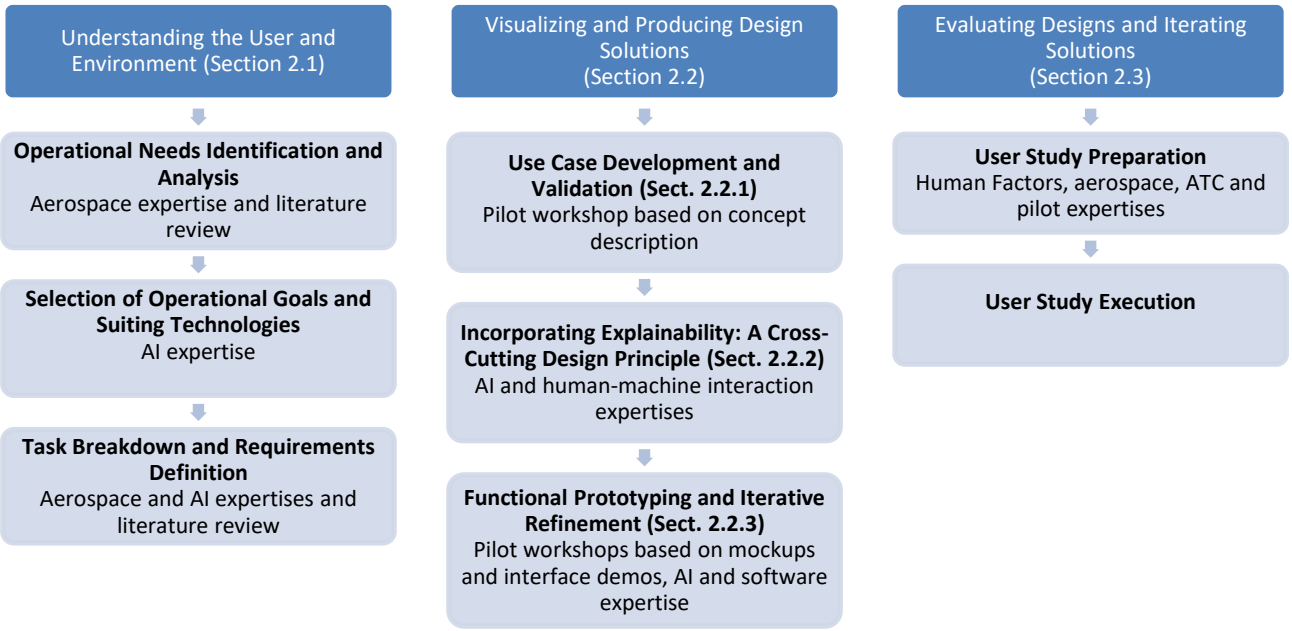


Figure 1 – Overview of the followed HCD process

2.1 Understanding the User and Environment

Understanding the user and the operating environment is critical for ensuring that design solutions align with user needs and operational constraints. This phase involved gaining a comprehensive understanding of the pilot's responsibilities, the challenges posed by the aviation environment, and the tasks required to accomplish operational goals. The following activities were undertaken as part of this phase:

Operational Needs Identification and Analysis Six operational goals were identified by aviation system subject matter experts, based on expertise and literature review [5, 6, 7], as representative of flight crew responsibilities during flight operations:

1. Monitor and manage aircrew, passengers, and cargo,
2. Manage the Flight Strategy,
3. Execute the flight,
4. Monitor aircraft systems and failures,
5. Monitor the external environment, and
6. Perform external communications.

Selection of Operational Goals and Suiting Technologies Among these goals, *Manage the Flight Strategy* and *Perform External Communications* were selected for the CAIRDA prototype. These goals were prioritized due to their alignment with technological capabilities enabled by AI, their potential to reduce workload, and their impact on operational efficiency and safety. For *Manage the Flight Strategy*, automated reasoning was identified as a valuable tool to assist pilots in evaluating and updating flight strategies in response to real-time changes in weather, airspace constraints, and aircraft performance. For *Perform External Communications*, Speech-to-Text (STT) technology was identified as an enabler to automate and streamline the interpretation and management of verbal ATC instructions.

Task Breakdown and Requirements Definition Subtasks were first defined for each selected operational goal based on domain expertise and established aviation guidelines provided by the International Civil Aviation Organization (ICAO) [8, 9, 10], the Federal Aviation Administration (FAA) [11, 12], and other relevant sources [13, 14, 15].

Tasks related to *Manage the Flight Strategy* included (A) assessing the need for a new flight strategy, (B) defining an updated flight strategy, and (C) evaluating computed flight profiles. For *Perform External Communications*, tasks included (A) interpreting ATC communications, (B) clarifying instructions, and (C) relaying information to other stakeholders.

To support these tasks, the interactions between the pilot and the digital assistant were defined using the Goal-Directed Task Analysis (GDTA) framework proposed by Endsley [16]. GDTA provides a structured approach for identifying operator goals, the decisions required to achieve those goals, and the information requirements needed to support those decisions in dynamic environments. For each operational domain, pilot goals were identified and used to derive the associated tasks, decision points, and information requirements. Within *Manage the Flight Strategy*, pilot goals focused on maintaining awareness of risks affecting the current flight strategy, identifying appropriate mitigation actions, and understanding the rationale behind potential changes to the flight plan. Within *Perform External Communications*, pilot goals focused on accurately understanding ATC instructions, verifying the correctness of communicated information, and maintaining awareness of communication-related actions. The resulting goals, tasks, decisions, and information requirements informed the design of the system functionalities and user interfaces, ensuring that the prototype addressed operationally relevant pilot needs while supporting effective decision-making.

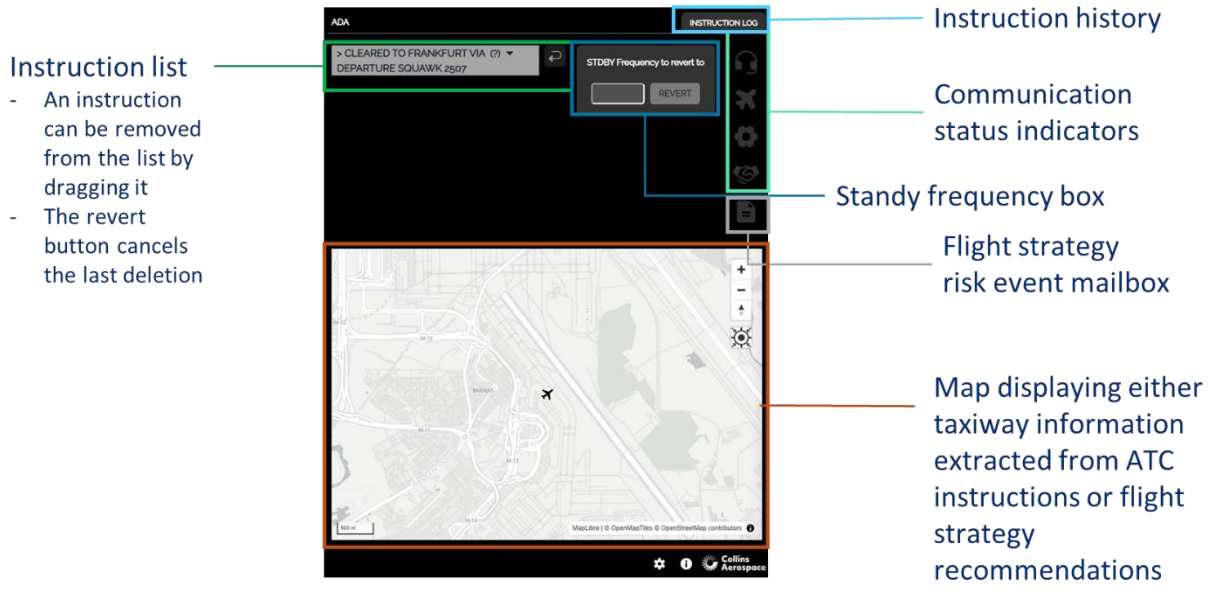


Figure 2 – Overview of the CAIRDA user interface
Map provided by MapLibre, ©OpenMapTiles ©OpenStreetMap

2.2 Visualizing and Producing Design Solutions

The first design and prototyping phase focused on translating the operational goals and user requirements into tangible system capabilities. This phase prioritized collaboration with pilots and subject matter experts to ensure the assistant's functionality, user interface, and behaviour aligned with operational needs. Based on the sub-goals and tasks defined in phase Understanding the User and Environment (2.1), two core functionalities evolved over multiple iterations to address the operational challenges identified during the initial research phase. The descriptions below correspond to the final concept, shaped through iterative prototyping, feedback integration, and collaboration with pilots and, aviation and human-machine interaction subject matter experts. The final user interface is shown in Figure 2.

Manage the Flight Strategy: Functionality Definition This functionality supports the sub-goals and tasks associated with identifying risks to the current flight strategy and assisting pilots in evaluating and updating the flight strategy to mitigate those risks. The CAIRDA system monitors internal and external triggers—such as system failures or meteorological events—that could impact the current flight strategy. When such triggers are detected, the system provides pilots with recommendations to mitigate risks and maintain operational safety. The system offers two types of flight strategy recommendations:

- Alternative Airports: When the situation requires the aircraft to land as soon as possible, the system identifies and suggests alternative airports. These recommendations are prioritized based on factors such as runway length, fire and hospital facilities, and distance from the aircraft's current position.
- Trajectory Adjustments: When a segment of the current flight trajectory is expected to encounter a meteorological event (e.g., turbulence, icing), the system recommends changes in speed or altitude to avoid the risk (see Figure 3).

To ensure pilots can act on these recommendations effectively, the system provides visual information on a flight map, highlighting affected trajectory segments and alternative airports alongside relevant data, such as runway length, METAR reports, and fuel considerations. This approach allows pilots to quickly comprehend the situation and evaluate the rationale behind the recommendations while maintaining control over their decision-making process.



Figure 3 – (Left) Map showing a representation of a meteorological event with the part of the trajectory which is at risk highlighted;
(2) Associated recommendation
Map provided by MapLibre, ©OpenMapTiles ©OpenStreetMap

Perform External Communications: Functionality Definition This functionality supports the sub-goals and tasks related to ensuring the accurate understanding and handling of ATC communications. The CAIRDA system automates ATC transcription and processes pilot readbacks to identify and validate instructions. Only matched instructions and elements are displayed in the instruction list, reducing the likelihood of miscommunication and enabling pilots to focus on validated information. To further support pilots in handling ATC communications, the system automates frequency settings by detecting and automatically updating the communication management system with new frequencies included in ATC instructions. To address pilot preferences for flexibility, the

system also allows frequencies to be reverted if necessary. Additionally, pilots are provided with multiple ways to visualize ATC information, including as text in the instruction list, on an airport map for navigation, and on a flight profile. These options were designed to accommodate diverse pilot preferences while minimizing cognitive load during high-workload flight phases.

2.2.1 Use Case Development and Validation

The first step in the **Visualizing and Producing Design Solutions** phase was the presentation of initial concepts considered for the CAIRDA system to 5 pilots during a dedicated workshop in March 2024. These concepts were developed based on the sub-goals and tasks identified in the phase Understanding the User and Environment (2.1), aiming to address the operational challenges associated with *Manage the Flight Strategy* and *Perform External Communications*. No implementation of the concepts had been completed at that stage; instead, the workshop focused on presenting the drafted concepts to pilots in order to gather early user feedback and validate a positive design direction.

The workshop started with a deep dive on each concept, including a comparison between current two-crew operations and the envisioned role of CAIRDA in advanced flight operations like for example Reduced Crew Operations (RCO). For each functionality, a hypothetical scenario was presented to demonstrate how the assistant could support pilots in RCO environments. Following their presentation, a series of questions was posed to the pilots to validate and refine the concepts. The following sections summarize the presented concepts, example scenarios, and key questions discussed during the workshop.

Manage the Flight Strategy: Concept Refinement The concept for the functionality focused on supporting pilots in reassessing and updating their flight strategies, in response to significant changes in the flight context. In current operations, pilots need to continuously monitor internal and external information from various sources (e.g., weather updates, airspace status, aircraft system status) to collaboratively assess and update the flight strategy. Both the pilot flying and pilot monitoring contribute to identifying significant changes to the reference flight context and deciding on necessary updates, ensuring mutual oversight and mitigating risks associated with high workload, fatigue, or stress. In RCO, the CAIRDA system was envisioned to assist a single pilot by continuously analysing internal and external inputs (e.g., METAR, NOTAM, and SIGMET reports, as well as aircraft parameters) to detect risks associated with the current flight strategy. The system would alert the pilot if a significant risk was identified and provide recommendations for mitigating that risk, such as suggesting trajectory adjustments or identifying suitable alternative airports. To ensure transparency, the system would also offer explainability upon request, allowing the pilot to review the reasoning behind its recommendations.

Feedback from pilots highlighted the importance of providing clear and actionable risk assessments and recommendations while avoiding information overload. Pilots pointed out the need to prioritize critical factors such as runway length, weather conditions, and fuel availability when suggesting alternative airports. They also stressed the importance of considering available airport facilities, such as parking locations, airstairs, and passenger services, to ensure the practicality of recommendations. Specific elements from external reports, such as turbulence and runway closures, were identified as having a high impact on decision-making and were considered essential for inclusion in the system's risk assessments.

Perform External Communications: Concept Refinement The concept for the functionality focused on addressing the challenges of managing ATC communications in RCO by automating key tasks, reducing pilot workload, and increasing communication accuracy. In today’s flight operations, the pilot monitoring and pilot flying collaborate to manage ATC communications, with tasks such as listening to ATC instructions, performing readbacks, inputting data into avionics, and monitoring changes handled manually. In future flight operations like RCO, the CAIRDA system was envisioned to assist a single pilot by transcribing ATC instructions and pilot readbacks, and matching the two to only display the final verified instructions in an organized instruction list. The system would also automate tasks like detecting and setting the standby ATC frequency, while ensuring pilots retained control over the process through a simple confirmation step.

Pilot feedback emphasized the importance of providing clear and concise communication outputs

and ensuring that workflows remain simple and easy to understand, particularly during high-workload scenarios. For the automatic frequency setting feature, pilots reached a consensus that an automatic frequency setting by the assistant was preferable, provided that the ability to confirm the change would be retained before the action was implemented. Regarding the display of ATC communications, there were divergent views on whether real-time transcription or only the final verified output was more useful, highlighting the need for further investigation of pilot preferences and operational requirements in future iterations.

2.2.2 Incorporating Explainability: A Cross-Cutting Design Principle

Explainability was treated as a fundamental design principle throughout the development of the CAIRDA prototype, ensuring that pilots could understand and trust the system's behaviour in a safety-critical environment. The integration of explainability was informed by a multidisciplinary approach and iterative development, guided by insights from an internal explainability workshop and feedback from pilots.

Development of the Explainability Layer

In January 2025, the development team, which included experts in aerospace, human-robot interaction, human factors, and artificial intelligence, held a dedicated explainability workshop to establish the conceptual framework for integrating explainability into the CAIRDA prototype. This workshop set out three key objectives: first, to enable pilots to understand the state of the system (e.g., whether it was listening, processing, or presenting results); second, to clarify the reasoning behind system outputs and recommendations by identifying the contributing factors; and third, to provide pilots with access to explanations at varying levels of detail, allowing them to explore additional information only when necessary, avoiding cognitive overload.

The outcomes of the workshop included the design of an icon menu to indicate system states and the creation of two multi-level explainability frameworks tailored to the prototype's two core functionalities: *Manage the Flight Strategy* and *Perform External Communications*. These frameworks were further refined during pilot workshops in February and March 2025.

Explainability as a Multi-Level Framework

The explainability workshop established a multi-level framework to provide pilots with flexible access to information. This framework supported both functional transparency and operational efficiency by balancing detailed reasoning with the need to avoid cognitive overload.

Manage the Flight Strategy: Explainability Concept The multi-level framework was designed to indicate events and risks in a structured and intuitive way. The normal state of the interface during cruise displayed a flight profile with relevant ATC instructions (e.g., related to altitude), alongside the icon menu. When a new event occurred but did not trigger a risk, a mailbox icon displayed a green check mark (see Figure 4, element (1)) along with the number of events in the mailbox. If any event triggered a risk, the flight profile was replaced by a map, which displayed information about affected trajectory segments or suitable alternate airports. In such cases, the mailbox icon displayed a yellow or red exclamation mark (see Figure 4, elements (2, 3)) depending on the severity of the risk, along with the number of events. This dynamic design ensured that pilots could intuitively monitor both normal and risk-triggering events and respond accordingly.

While the system's underlying automated reasoning capabilities could provide detailed traces of its computations, the development team deliberately limited the level of real-time detail provided to avoid overburdening pilots. Instead, the type of information used by the system (e.g., rules implemented in the automated reasoning component) and the overall functioning of the system were documented in the user manual to support pilot training. In the future, it is envisioned that computation traces could be logged to assist with traceability efforts. The system's automated reasoning design also provides a strong foundation for potential certification, as it can explicitly outline the steps in its reasoning process and explain how it reached a specific result, an advantage over machine learning-based systems.

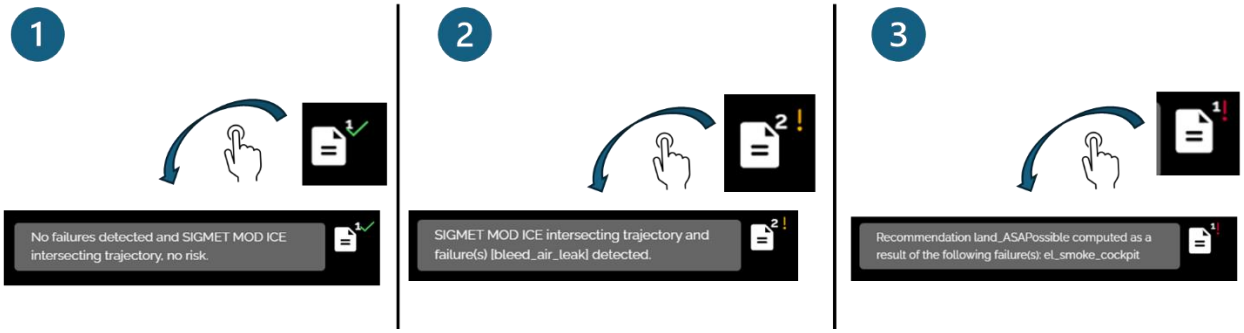


Figure 4 – Event mailbox and risk explanation for (1) an event observed without identified risk; (2) a combination of 2 events posing a risk; (3) an event posing a risk which requires to land

Perform External Communications: Explainability Concept Three levels of explainability were developed:

1. The first level relied on visual indicators such as a yellow or red handshake icon and question marks in the instruction list to highlight discrepancies (see Figure 5, elements (1) and (3)).
2. The second level allowed pilots to click on the handshake icon to view the specific mismatching elements in both the ATCO instructions and pilot readbacks (see Figure 5, element (4)).
3. The third level enabled the pilot to click on the mismatched elements to access the raw speech-to-text transcription for both the ATCO and pilot, offering a deeper level of transparency (see Figure 5, element (5)).

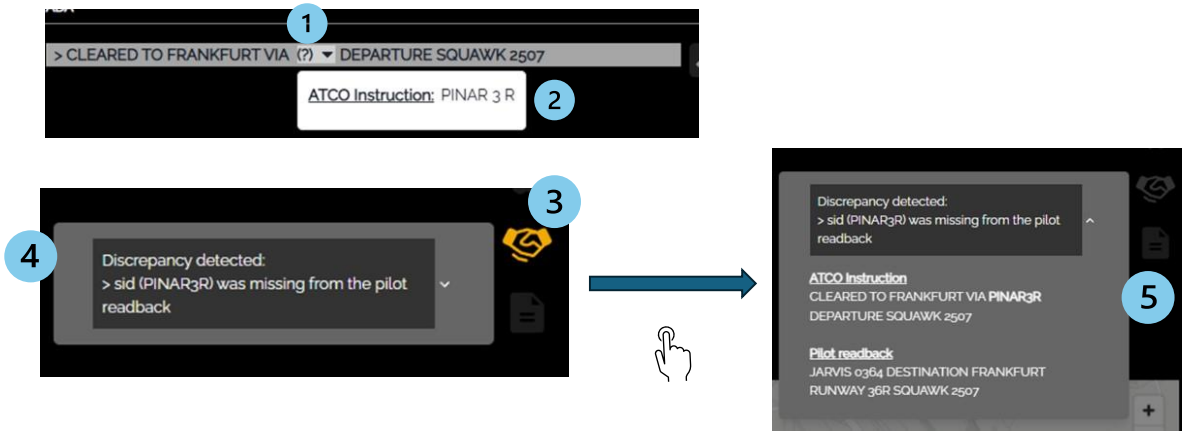


Figure 5 – (1) Question mark symbol indicating a missing SID in the pilot's readback; (2) ATC instruction value displayed upon clicking (1), allowing insertion into the instruction; (3) Yellow handshake indicating a discrepancy; (4) Mismatching values highlighted; (5) Expanded view of (4) showing the ATC instruction and pilot readback transcriptions.

By embedding explainability as a cross-cutting principle, the CAIRDA prototype was designed to enhance pilot trust, usability, and situation awareness. The integration of a multi-level explainability framework ensured that pilots could access information with varying levels of detail, depending on the operational context, without being overwhelmed by unnecessary complexity. The refinements made during the iterative development process—such as improved risk explanations, dynamic event displays, and enhanced control over discrepancies—demonstrated the importance of aligning system behaviour with pilot needs. Ultimately, explainability played a critical role in making the CAIRDA prototype both functional and interpretable, ensuring its suitability for use in a safety-critical aviation environment.

2.2.3 Functional Prototyping and Iterative Refinement

Following the validation of the initial concepts during the March 2024 pilot workshop, the design team focused on implementing a first version of the CAIRDA prototype and refining it through iterative feedback. By early 2025, the prototype included core functionalities for both *Manage the Flight Strategy* and *Perform External Communications*, which were evaluated through a series of pilot workshops and human-in-the-loop (HITL) testing. This phase was critical for refining the system's human-machine interface (HMI), interaction workflows, and AI algorithms in preparation for higher-fidelity evaluations.

The design team conducted two pilot workshops in February and March 2025: one with two US-based pilots (online) and another with three German pilots (in person). During these workshops, pilots were shown screenshots of the developed prototype and mockups of partially implemented features (e.g., the explainability layer). Pilots were asked for feedback on the system's usability and design, as well as to respond to targeted questions about specific elements of the interface and workflows. In parallel, HITL testing in an ATR flight simulator was conducted with seven pilots, providing additional insights into the system's usability and performance in operationally representative environments. A summary of the implemented refinements can be found in Table 1.

	Feature	Initial Design	Refinement Implemented
Manage the Flight Strategy	Alternative Airport Prioritization	Basic prioritization logic for selecting alternative airports in emergencies.	Introduced weights and priorities for alternative airports based on factors like fire facilities, distance, and hospital availability. Prioritization logic was tailored to specific scenarios.
	Alternative Airport Display	Alternative airports displayed without clear priority or critical information.	Limited alternative airports displayed on the map to three. Added critical information for each airport: runway length, METAR, elevation, approaches, fuel remaining on landing, and fuel to dump before landing.
	Explainability - Risk Descriptions	Risk descriptions were lengthy and potentially overwhelming.	Simplified risk descriptions to ensure clarity and reduce cognitive load. Replaced full SIGMET text with only the most relevant information (e.g., icing conditions). Retained failure and SIGMET combination explanations for transparency and trust-building.
Perform External Communications	ATC Transcription Display	Only final processed ATC instructions displayed.	Added real-time transcription for pilots to compare text with notes and catch errors, such as incorrect taxiways.
	Frequency Setting Workflow	Two modes: manual confirmation or automatic frequency setting with a pop-up.	Introduced frequency box: automatically sets standby frequency with a revert option and a pop-up notification.
	Explainability - Discrepancy Resolution	Limited control over resolving discrepancies in ATC communication.	Added a feature allowing pilots to click on discrepancy symbols in the instruction list and manually select the correct ATCO value, enabling real-time resolution and increased confidence in outputs.
Both	Visual Clarity of HMI Elements	Text and visual elements were not optimized for readability or usability.	Adjusted size, colour, and placement of text and visual elements. Updated holding points and trajectory risk icons.

Table 1 – Evolution of selected features through iterative refinements

Manage the Flight Strategy: Iterative Refinements on Functionality, Usability & Explainability

The *Manage the Flight Strategy* functionality saw significant refinements to improve the prioritization of recommendations, display elements, and the assistant's explainability:

- Map and Airport Displays: Initially, all waypoints were displayed on the map. After the US-pilot workshop, this was updated to display only waypoints affected by meteorological events or the next waypoints on the flight path. For alternative airports, feedback from both workshops established a priority list of critical information to display: runway length, METAR, elevation, available approaches, fuel remaining at landing, and fuel to dump before landing. To reduce cognitive overload, the number of alternative airports displayed was limited to three.
- Algorithm Refinements: Pilot feedback was instrumental in refining the assistant's AI reasoning for selecting alternative airports. After the workshops, weights and priorities were introduced to account for specific factors, such as hospital and fire facilities. For example, in use cases involving electric smoke, fire facilities were prioritized over distance and hospital availability, while for an antiskid fault combined with engine fire, distance was prioritized over fire facilities and hospital availability due to increased landing distance requirements. Discussions with pilots and SMEs clarified that the system would always propose solutions in accordance with flight manuals and legal requirements, but pilots retained the authority to deviate from these suggestions in emergency situations.
- Alerting Behaviours: The initial system used a side menu with an icon to signal detected risks. Feedback from the German-pilot workshop highlighted the potential for such alerts to go unnoticed during high-workload scenarios. As a result, a more prominent alerting mechanism was introduced, where the screen blinks red until the pilot acknowledges the alert by clicking on it.
- Text Display: Pilots provided input on textual displays, such as using Mach instead of knots for cruise speed recommendations and changing the term "recommend" to "consider" emphasizing pilot authority.
- Explainability: The phrasing of the risk descriptions was simplified to ensure clarity and reduce cognitive load. Additionally, instead of displaying the entire SIGMET text, the system was updated to extract and display only the most relevant information, such as icing conditions. Although pilots suggested that it was unnecessary to explicitly highlight that a failure combined with a SIGMET triggered a detected risk (since they were already aware of failures in their aircraft), the development team retained this information. This decision reflected the importance of endowing the assistant with an explainability behaviour, providing transparency into its reasoning process and building trust in its recommendations.

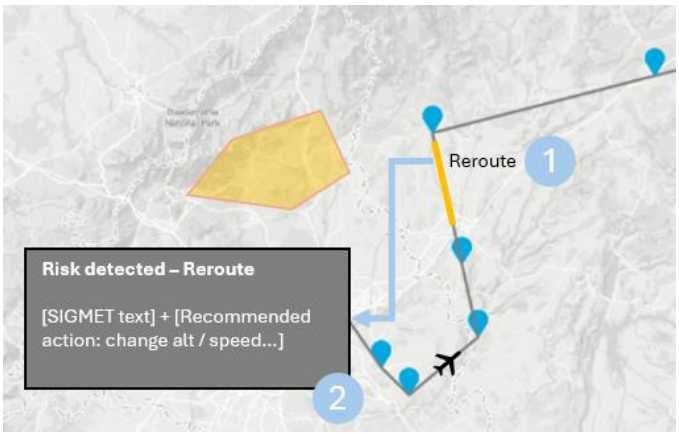


Figure 6 – PowerPoint mockup presented during a pilot workshop for the trajectory adjustment feature of *Manage the Flight Strategy*
Map provided by MapLibre, ©OpenMapTiles ©OpenStreetMap

Perform External Communications: Iterative Refinements on Functionality, Usability & Explainability

The *Perform External Communications* functionality underwent several critical updates based on feedback from the pilot workshops and HITL testing:

- Comparison Between ATC and Pilot Speech: The system initially compared ATC instructions with pilot readbacks to display only verified information. However, during HITL testing, it became clear that pilot readbacks often lacked the structured format of ATC instructions, making the semantic analysis prone to errors. This prompted a complete re-implementation of the pilot grammar and a refactoring of the comparison algorithm to improve accuracy.
- Real-Time Display of ATC Speech: During the initial use case workshop, pilots expressed a preference for displaying only the final processed ATC instruction to avoid potential communication errors. However, feedback from the US-pilot workshop highlighted the value of real-time transcription for identifying errors, such as incorrect taxiway names, by comparing text with notes. Based on this feedback, the system was updated to include real-time transcription alongside the final verified outputs.
- Frequency Setting Workflow: The initial design included two modes for frequency setting: one requiring manual confirmation and another where frequencies were set automatically with a notification. Feedback from both workshops led to a consensus on automatically setting the frequency but with the ability to revert to the previous value if needed. The final design introduced a frequency box, where a detected frequency is automatically set as the standby frequency, with a pop-up notification and a “revert” option to restore the previous frequency.
- Instruction List and History Features: Instructions were initially displayed in chronological order with the oldest at the top. Pilots recommended reversing the order to show the most recent instructions at the top, which was implemented in subsequent iterations. Pilots also validated the auto-cleanup feature, which removes instructions older than two flight phases or 25 minutes, and acknowledged that no confirmation was needed when pilots were deleting instructions themselves, as they remain accessible in the history log. A replay feature was added following the German-pilot workshop to allow pilots to revisit readbacks, especially in busy environments where ATC clarifications might be missed.
- Explainability: To provide pilots with more control over discrepancies, the German-pilot workshop led to the addition of a feature that allowed pilots to click on the discrepancy symbol in the instruction list and manually select the ATCO value if they determined it to be correct (see Figure 5, element (2)). This refinement addressed the need for pilots to resolve discrepancies in real time and improved their confidence in the system’s outputs.

HITL Testing and Transition to Higher-Fidelity Evaluations

The HITL testing in the ATR flight simulator, conducted with seven pilots, played a key role in refining the system’s usability and preparing for higher-fidelity evaluations. Feedback from this testing led to improvements in the size and colour of HMI elements, map zoom functionality, and the addition of an aircraft centring feature on the map. These refinements helped address usability concerns, emphasizing the importance of ensuring a robust and intuitive interface so that pilots could focus on evaluating the core concepts and functionality of the system effectively.

The iterative design process also benefited from HITL testing by identifying challenges related to pilot training on the system. As a result, the team developed a more comprehensive user manual and implemented a tutorial interface to accelerate pilot familiarization. These enhancements were critical to ensuring that future evaluations would focus on the assistant’s core features and decision-making support, rather than being limited by basic usability issues.

The combined findings from the pilot workshops and HITL testing and the implementations of the refinements ensured the prototype was ready for the high-fidelity evaluations described in the next section.

2.3 Evaluating Designs and Iterating Solutions

Evaluation and iteration were critical to improving the effectiveness, usability, and operational alignment of the CAIRDA prototype. This phase focused on preparing the system for a higher fidelity evaluation by addressing remaining design issues, validating the realism of the test scenarios, and implementing final refinements. An important activity during this phase was a pilot instructor workshop, conducted with two commercial pilot instructors. During this workshop, the developers walked the instructors through the system's scenarios, behaviour, and interface, collecting feedback to refine the prototype further. This collaborative session focused on ensuring the scenarios were realistic and operationally relevant while also fine-tuning the system's usability. The feedback from this workshop ensured that the CAIRDA prototype was ready for the final user study, which served as the summative evaluation of the system. The methodology of the final user study is described in Section 2.4 while the results and findings are presented in Section 3.

2.4 Methodology of the Final User Study

The final user study was conducted to evaluate the CAIRDA prototype as part of the Evaluating Designs and Iterating Solutions phase of the HCD process. The study assessed whether CAIRDA effectively supported pilots in managing flight strategies and performing external communications in a simulated operational environment.

The study utilized the iSIM, a flat-panel cockpit simulator by the Institute of Flight Guidance of the German Aerospace Center, which replicates an Airbus A321 flight deck, allowing for the integration of the CAIRDA system. 21 Airbus type-rated pilots participated, with an average age of 40 years and an average flight experience of 7,657 hours. The study used a 2 (flight: reference vs. solution) x 2 (situations: taxi & takeoff, cruise) mixed design, with one between-subjects factor (transparency) and two within-subject factors (flight and situation).

The experiments were conducted with one pilot participating in each run. Each pilot completed a reference flight without CAIRDA support and a solution flight with CAIRDA assistance. The user study focused on two operational flight scenarios, each designed to evaluate one of the two operational goals selected for the CAIRDA prototype:

1. Taxi and Takeoff: This scenario tested the CAIRDA's ability to support the operational goal of *Perform External Communications*. Pilots navigated from the taxi stand to the runway and performed takeoff. In the solution flight, CAIRDA assisted with transcription, instruction verification, and visual guidance.
2. Cruise: This scenario was designed to assess the CAIRDA's support for the operational goal of *Manage the Flight Strategy*. Pilots managed weather and technical disruptions requiring flight risk assessment. CAIRDA supported decision-making by detecting risks, cross-referencing technical and environmental data, and providing clear recommendations.

Data collection methods included subjective questionnaires (e.g., DLR Workload Assessment Tool [17], SATI [18], System Usability Scale [19], SHAPE Teamwork [20]), eye-tracking for attention analysis, and behavioural detection tasks to evaluate pilots' responses to testable events. Statistical analysis employed multi-level modelling to assess the effects of flight condition, transparency, and situation context on key metrics such as workload, trust, and situation awareness.

3. Final User Study Results

To evaluate the CAIRDA prototype's impact on pilot task performance, the final user study assessed its ability to support pilots in *managing the flight strategy* and *performing external communications*. The results are organized into the following dimensions: operational feasibility, workload, trust and teaming, situation awareness, usability, and safety.

3.1 Operational Feasibility of CAIRDA Functionalities

The operational feasibility of CAIRDA was assessed by comparing the task duration between reference and solution flights for both functionalities. Task duration was measured from the point at which the experimenter provided the necessary instructions and information for a given situation until the pilots completed their tasks. Across all scenarios, no significant differences in task duration were found ($p = 0.709$), confirming that pilots completed their tasks in a comparable timeframe regardless

of whether CAIRDA was used. This indicates that the system is operationally feasible, as it does not impose additional delays on task execution while supporting pilots in managing flight strategy and performing external communications.

For *Perform External Communications*, additional analyses were conducted to assess communication efficiency on the level of ATC instructions. As shown in Figure 7, CAIRDA significantly improved communication efficiency by reducing the number of communication turns, minimizing pilot queries, and decreasing the time needed to complete full readback of an ATC instruction. These findings underline the system's ability to streamline communication processes, contributing to its feasibility in operational use.

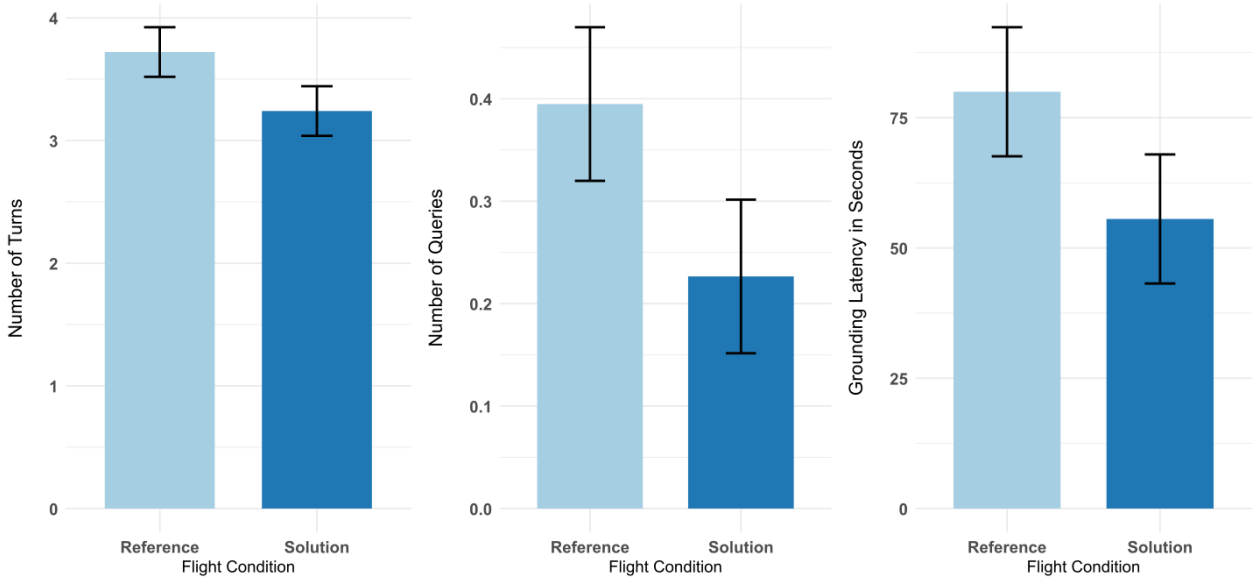


Figure 7 – Communicative efficiency for ATC cockpit instructions without CAIRDA (reference) and with CAIRDA (solution).

3.2 Impact of CAIRDA on Pilot Workload

The impact of CAIRDA on pilot workload was evaluated using the DLR Workload Assessment Tool [17], which measures workload across eight dimensions on a scale from 0 (underload) to 200 (overload), with 100 indicating optimal workload. The results of a mixed linear model showed that CAIRDA effectively reduced overall workload for both *Manage the Flight Strategy* and *Perform External Communications* during solution flights compared to reference flights.

For *Manage the Flight Strategy*, workload decreased from M = 90.3 to M = 84.4 (X^2 (1) = 4.318, p = 0.038), while for *Perform External Communications*, workload dropped from M = 104.3 to M = 91.3 (X^2 (1) = 13.371, p < 0.001). These reductions indicate that CAIRDA provided effective task support by alleviating the cognitive demands of both functionalities. For *Perform External Communications*, the reductions were particularly notable in the information acquisition (estimate = -10.571) and effort (estimate = -8.952) dimensions, suggesting that the transcription and instruction verification features helped pilots focus on higher-priority tasks.

Overall, the results support the conclusion that CAIRDA reduced workload and provided effective assistance. A breakdown of workload ratings is shown in Figure 8.

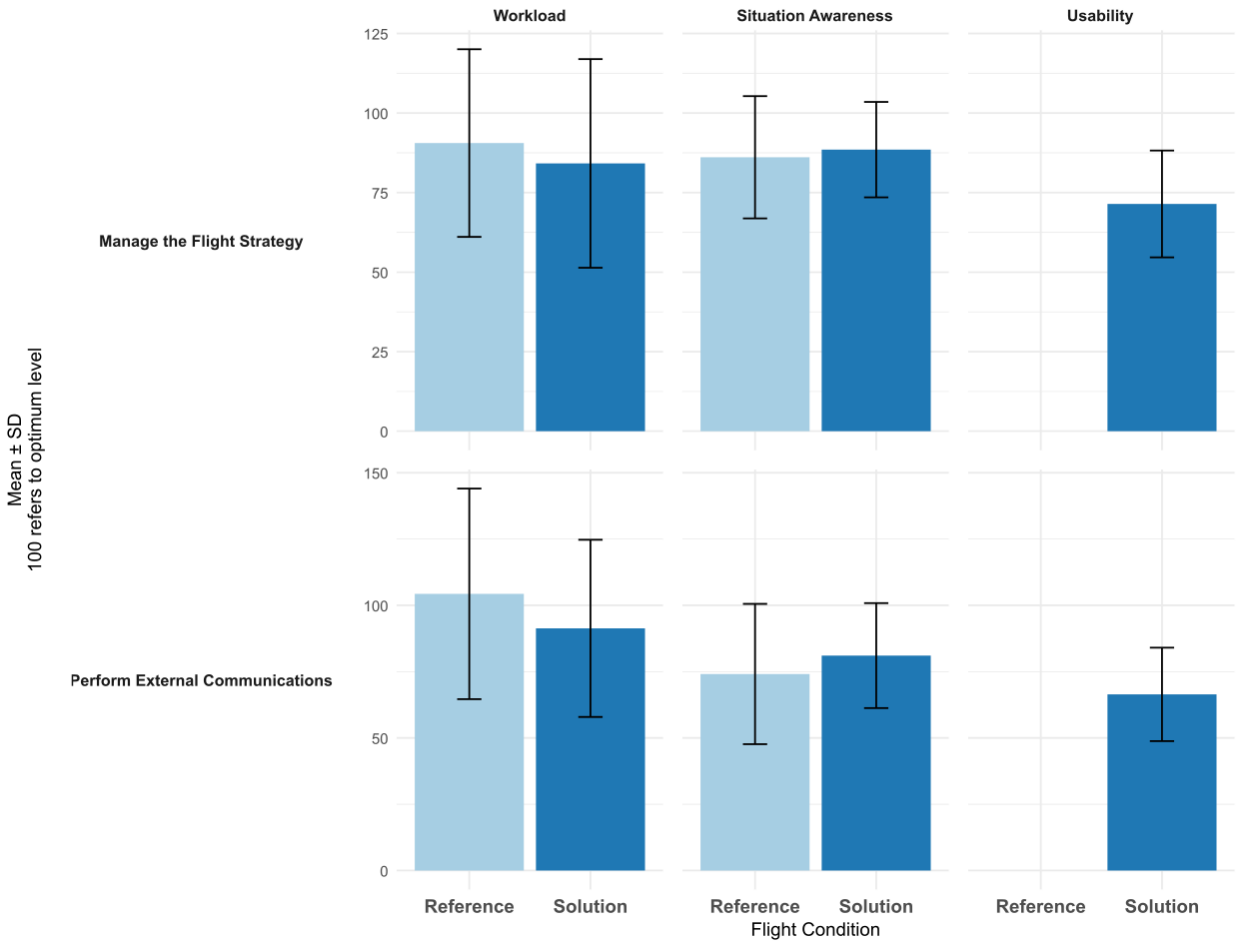


Figure 8 – Workload, Situation Awareness and Usability ratings for *Manage the Flight Strategy* and *Perform External Communications* operational goals under the reference (with CAIRDA) and solution (without CAIRDA) conditions

3.3 Trust and Human-AI Teaming in CAIRDA Operations

The CAIRDA system's impact on trust and human-AI teamwork was assessed using the SATI questionnaire (for trust) [18] and the SHAPE Teamwork questionnaire (for teamwork) [18]. The results show that pilots had a positive perception of CAIRDA as a supportive and trustworthy teammate in both *Manage the Flight Strategy* and *Perform External Communications*.

For trust, pilots rated CAIRDA positively on a 7-point Likert scale for both functionalities (M = 4.02, SD = 1.47 for *Manage the Flight Strategy*; M = 4.02, SD = 1.53 for *Perform External Communications*)

For teamwork, 19 out of 21 pilots acknowledged clear task responsibility across both functionalities. Additionally, 12 pilots reported that CAIRDA effectively supported task prioritization for *Manage the Flight Strategy*, and 13 pilots noted a smooth flow of information for *Perform External Communications*. While these findings indicate that the system contributed to fostering human-AI collaboration for a majority of pilots, they also suggest room for further improvement to enhance the overall experience for all users.

3.4 Usability of the CAIRDA System

The usability of the CAIRDA system was evaluated using the System Usability Scale (SUS) [19], which measures user satisfaction and ease of use on a scale of 0 to 100, with higher scores indicating better usability. Overall, usability was rated as good.

For both *Manage the Flight Strategy* and *Perform External Communications*, usability scores were high during solution flights, with average ratings of M = 71.43 (SD = 16.78) and M = 66.43 (SD = 17.62), respectively. A breakdown of usability ratings is shown in Figure 8. During reference flights no assistance system was present and thus no usability ratings have been administered.

3.5 Effects of CAIRDA on Situation Awareness

The impact of CAIRDA on situation awareness was evaluated through subjective ratings, eye-tracking analysis, and behavioural detection tasks. Overall, the results demonstrated that pilots maintained their situation awareness in the cockpit environment when using CAIRDA, although attention shifted toward the system during specific flight phases.

Subjective ratings of situation awareness showed no significant differences between solution and reference flights for either functionality, with scores averaging M = 87.85 vs. M = 86.75 ($p = 0.767$) for *Manage the Flight Strategy*, and M = 81.05 vs. M = 74.10 ($p = 0.602$) for *Perform External Communications*. A breakdown of situation awareness ratings is shown in Figure 8.

Eye-tracking analysis revealed significant shifts in attention patterns during solution flights ($p < 0.001$). For *Manage the Flight Strategy*, pilots directed more attention to CAIRDA during the cruise phase ($p < 0.001$), indicating active use of the system for decision-making. For *Perform External Communications*, pilots focused more on the CAIRDA interface ($p < 0.001$) and less on external views ($p < 0.001$) and the pedestal ($p = 0.015$), reducing "eyes-out" time during taxi operations.

Behavioural detection tasks were designed to assess pilots' ability to identify intentional procedural discrepancies (e.g., incorrect ATC frequencies or holding points) during taxi and takeoff, as well as failure events (e.g., weather radar failure or bleed leak) during the cruise phase. The results showed no significant differences in detection success between reference and solution flights, confirming that pilots were able to identify procedural discrepancies and failure events equally well in both conditions.

While CAIRDA supported cockpit-related situation awareness effectively, the observed shift in attention toward the system underscores the need for further refinements to ensure pilots maintain adequate external awareness during critical flight phases.

3.6 Maintaining Safety with CAIRDA

The user study confirmed that CAIRDA maintained the level of safety across all test scenarios, with no significant differences observed in pilots' ability to detect procedural discrepancies or failure events between reference and solution flights.

For *Manage the Flight Strategy*, CAIRDA provided operationally feasible recommendations adapted to the technical and environmental conditions of the aircraft. However, at least half of the pilots indicated that task prioritization during failure scenarios was unclear, particularly regarding whether to first address CAIRDA's recommendations or perform ECAM actions. Although this ambiguity did not result in unsafe situations, further refinement of procedures could help improve safety during time-critical scenarios.

For *Perform External Communications*, pilots directed more attention to the CAIRDA display and less to external views during taxi operations, particularly when addressing transcription problems or discrepancies between ATC instructions and pilot readbacks. While accessing these discrepancies helped pilots ensure correct instructions were displayed, it also reduced external situation awareness. Misrouting incidents were observed in both reference flights (three pilots) and solution flights (five pilots), with no significant difference between the two conditions ($p = 0.694$).

Overall, pilots highlighted that CAIRDA was a valuable tool for maintaining safety, particularly in identifying and managing risks. However, they noted areas for improvement, particularly the need for refining task prioritization during failure management and ensuring that the system's guidance aligns seamlessly with existing operational procedures. These refinements are necessary to support seamless integration into operational workflows and to maintain the high safety standards required in aviation.

The user study confirmed that CAIRDA effectively supports pilots in managing the flight strategy and performing external communications, demonstrating its operational feasibility, reducing workload, and fostering trust and usability without compromising safety. While the system successfully maintained cockpit-related situation awareness, further refinements are needed to improve attention distribution, task prioritization, and integration into existing workflows. These results provide a strong foundation for advancing CAIRDA toward higher technology readiness levels while addressing areas for improvement identified during this user study.

4. Discussion

The final user study of the CAIRDA prototype provided valuable insights into the application of the HCD process for developing AI-based systems in safety-critical aviation environments. By aligning system functionalities with real-world operational needs, embedding explainability as a core principle, and iteratively refining the design based on user feedback, the project demonstrated how HCD can address the dual goals of enhancing operational performance and ensuring safety. This section reflects on the study's findings in terms of their implications for HCD, human-AI collaboration, and the future design of cockpit systems.

4.1 Aligning System Design with Operational Goals

The results confirm that the HCD process successfully translated the operational goals of *Manage the Flight Strategy* and *Perform External Communications* into effective system functionalities. The selection of these goals reflected critical pain points in pilot workload, communication, and decision-making, which were validated through engagement with pilots and subject matter experts.

The *Manage the Flight Strategy* functionality supported pilots in navigating dynamic and complex flight contexts by aggregating and analysing multiple data streams to assess risks, provide actionable recommendations, and offer clear explanations for its guidance. These functionalities demonstrate how AI can complement pilot capabilities by automating routine tasks and assisting with decision-making in more complex scenarios. Similarly, the *Perform External Communications* functionality addressed challenges in ATC communication by automating the transcription of instructions, validating pilot readbacks, and presenting confirmed instructions in clear and actionable formats. The deployment of features such as automatic frequency updates and validated instruction displays on maps and timelines demonstrated the potential of AI to reduce workload and streamline communication workflows.

4.2 Integrating Divergent Operational Goals into a Unified Design

The development of CAIRDA required balancing two distinct operational goals within a unified design framework. To achieve this, the team employed a dual approach: the *Same-Path Approach* ensured consistency, while the *Different-Path Approach* addressed unique operational requirements.

Same-Path Approach A corporate design style guide was ESTABLISHED early, defining a cohesive colour scheme, typography, icon set, and layout philosophy. A design system was created to streamline UI elements, enabling consistency across functionalities. The human-centred design process ensured pilot involvement at every stage, allowing iterative feedback to shape the design from wireframes to prototypes. A shared *Explainability Layer Framework* was implemented for both goals, offering a consistent three-stage model: status indicators, concise explanations, and optional drill-downs to raw data. This approach supported trust and usability without overloading the interface.

Different-Path Approach Each operational goal required tailored design solutions. For *Manage the Flight Strategy*, the system identified risks, such as system failures or meteorological events, and provided pilots with recommendations, including alternative airports and trajectory adjustments, displayed on a flight map with key data like METAR reports and fuel considerations. This enabled quick risk assessment and informed decision-making. For *Perform External Communications*, the focus was on automating ATC transcription, validating pilot readbacks, and presenting matched instructions in clear formats. Features like automatic frequency updates and multiple visualization options—text-based lists, airport maps, and flight profiles—streamlined communication workflows while minimizing pilot cognitive load.

4.3 The Role of Explainability in Enhancing Trust and Usability

The explainability features of the CAIRDA prototype were designed to provide pilots with clear, actionable insights into the system’s outputs, aligning with HCD principles for transparency and trust. While the user study did not directly measure the impact of explainability, the positive results for trust and usability suggest that these features may have played a role in the system’s effectiveness.

For both *Manage the Flight Strategy* and *Perform External Communications*, the layered explainability features such as the yellow handshake to flag mismatches in ATC communications or the contextual explanations for flight strategy risks, were intended to help pilots understand and trust

the system's outputs. By providing on-demand access to deeper insights and justifications, the features allowed pilots to verify information at their discretion, without significantly more time consumption, potentially fostering confidence in the system while minimizing cognitive overload.

Although the study did not explicitly assess pilot interaction with these features, the trust and usability scores indicate that explainability likely contributed to the system's ability to support decision-making and operational performance. Future research could explore the specific role of explainability to better understand its direct impact on trust, usability, and situation awareness.

4.4 Balancing AI-Based Decision Support and Attention Management

The final user study revealed the challenge of ensuring that AI-based decision-support systems like CAIRDA enhance pilot performance without unintended consequences for attention management, particularly in safety-critical environments where situation awareness is paramount. While the CAIRDA prototype successfully reduced pilot workload and enhanced operational efficiency, the observed shift in attention towards the system during certain flight phases underscores the need for careful and thorough interface design. For example, during taxi and take-off, pilots focused more on the CAIRDA interface to review ATC instructions, which temporarily reduced their external visual attention.

This finding highlights a key challenge in designing AI-based systems: ensuring that automation aids pilots without distracting them from their primary tasks or compromising their situation awareness. Future iterations of CAIRDA should focus on refining the interface and interaction design to better balance system use with the need for external awareness, particularly during critical flight phases.

4.5 Ensuring Safe and Seamless Workflow Integration

The final user study confirmed that CAIRDA maintained safety across all test scenarios, with no significant differences observed in pilots' ability to detect procedural discrepancies or failure events between reference and solution flights. This suggests that the system's functionalities and explainability features supported pilots without introducing new risks.

The workshops also prompted a broader question about the evolution of cockpit procedures in response to new technologies. For example, during failure scenarios, pilots expressed uncertainty about whether to prioritize CAIRDA's recommendations or follow standard ECAM procedures. While this points to a need for clearer task prioritization mechanisms within the system, it also raises the question of whether established procedures may need to evolve to fully leverage the capabilities of AI-based systems like CAIRDA.

Rather than solely focusing on adapting new technologies to current procedures, the integration of AI systems presents an opportunity to revisit and potentially modernize traditional workflows. For instance, cockpit procedures could be re-evaluated to incorporate AI guidance as an integral component of decision-making, rather than as a secondary tool. Such an approach could help ensure more seamless collaboration between pilots and AI systems while maintaining clear task hierarchies and preserving safety. These findings emphasize the need for a holistic approach to integrating new technologies into the cockpit, considering not only the design of the systems themselves but also the broader operational context in which they are deployed.

4.6 Broader Implications for HCD in Safety-Critical Domains

This study highlights the strengths of HCD as a process for developing AI-based systems in safety-critical environments, such as aviation. By aligning system functionalities with operational needs and involving end-users throughout the design process, HCD ensures that systems are both usable and effective in supporting human operators in high-stakes contexts. At the same time, the study points to opportunities for refining HCD practices to address some of the unique challenges of these domains.

The study reinforces the importance of integrating explainability as a fundamental design consideration in AI-enabled systems. Explainability is critical for building trust and ensuring safe human-machine collaboration in environments where operators need to monitor and intervene in automated processes. The CAIRDA prototype's layered explainability features, such as the yellow handshake and contextual recommendations, illustrate how transparency can be embedded into the

design. However, future HCD efforts should seek to explicitly address explainability in a more systematic way, including the development of tools and metrics to measure its impact on trust, decision-making, and operational effectiveness.

Finally, while this study focused on adapting AI to existing workflows, the findings suggest that HCD could also play a role in shaping the development of new workflows that can better integrate AI systems. For example, pilot uncertainty about task prioritization during failure scenarios highlights the potential for AI systems to challenge traditional procedures. Rather than solely adapting AI systems to fit current workflows, HCD practitioners could work closely with operational stakeholders to explore how workflows might evolve to take full advantage of AI's capabilities while maintaining safety and supporting human operators. This collaborative approach could help ensure that technological advancements are thoughtfully integrated into safety-critical environments.

By addressing these broader implications, HCD can continue to support the development of AI systems that enhance safety, situation awareness, and operational efficiency in complex and demanding environments.

5. Challenges and Limitations

The development and validation of the CAIRDA prototype provided valuable insights into the strengths and weaknesses of applying HCD in the context of developing AI systems for safety-critical environments. However, several limitations and challenges emerged during different phases of the project, illuminating areas for improvement. These are discussed below.

5.1 Challenges and Limitations During User Research

Balancing Diverse Pilot Preferences Variability in pilot preferences and expectations posed a significant challenge throughout the user research and design phases, even within HCD framework. While HCD emphasizes incorporating user input, the diversity of pilot opinions complicated the development of universally acceptable solutions. Some pilots sought detailed system transparency and interactive features for greater operational understanding, while others preferred streamlined interfaces to minimize distractions. These differing needs required the development team to make trade-offs, often favouring simplicity to meet operational constraints, which may not have fully addressed all user needs. These challenges highlight the importance of further adapting the HCD approach to balance diverse user requirements with system usability. Exploring adaptive interfaces in future iterations could help align the system with varying pilot preferences while maintaining operational efficiency.

Limited Access to End Users for Research Access to pilots was limited due to scheduling and operational constraints, making it difficult to gather diverse and representative input during user research. This may have narrowed the range of perspectives and operational contexts considered, potentially influencing the design process. Feedback was also shaped by the operating philosophies of specific aircraft types, as pilots trained on Boeing and Airbus models often have differing expectations for automation and workflows.

This limitation highlights the importance of considering the diversity of operational philosophies when designing AI-based systems for safety-critical environments. Future HCD efforts should aim to involve pilots with experience across a range of platforms to ensure that design decisions are adaptable and inclusive, particularly when developing systems intended for use in diverse operational contexts.

5.2 Challenges and Limitations During Design and Prototyping

Balancing Iterative Design with Development Timelines The HCD process is built on the principle of iterative refinement, but real-world development timelines often impose limits on the number of iterations that can be conducted. In this project, internal deadlines required the team to prioritize certain design aspects over others, leaving limited time for deeper exploration of features such as confidence indicators or "why not" explanations for alternative recommendations. While the iterative cycles that were conducted were meaningful, this constraint highlights a key limitation of HCD in time-sensitive projects: the need to ensure that limited iterations effectively address the most critical design challenges.

Balancing Transparency with Usability One of the most significant challenges during the design phase was operationalizing explainability in a way that balanced transparency with usability. For example, for the *Manage the Flight Strategy* functionality, explainability was focused on providing clear rationales for recommendations, but further refinements are needed to include "why not" explanations. These challenges emphasize the difficulty of designing AI systems that enhance transparency while minimizing cognitive load. Similarly, the *Perform External Communications* feature required presenting sufficient information to allow pilots to understand mismatches in ATC communications (e.g., through the yellow handshake and question marks in the instruction list) without causing excessive "eyes-on-screen" time during high-workload flight phases.

5.3 Challenges and Limitations in Aligning Training and Operational Use

The distinction between training and operational use emerged as a critical area where the design process needs to account for how user feedback evolves over time. This project highlighted that pilots' initial impressions of a system during workshops or early evaluations may be influenced by their lack of familiarity with that system's features and functionalities. For example, a feature that is initially perceived as complex or not particularly usable may, with robust training and repeated use, become better understood and appreciated by pilots, especially given their high levels of skill and adaptability.

To address this, it may be helpful to design user study sessions that simulate varying levels of familiarity, from first-time exposure to more advanced operational scenarios. This would allow researchers to assess how user perceptions and preferences evolve over time, helping to distinguish between genuine usability issues and those that require additional training. Such an approach could help ensure that the design process supports both short-term usability and long-term operational effectiveness.

5.4 Challenges and Limitations in Addressing Future User Needs

Aerospace projects face a key challenge: the gap between the expertise and needs of test users during development and the evolving capabilities of future end-users. Pilots involved in validation are often experienced professionals, while later generations may be more familiar with modern technologies like digital assistants, adopting systems more intuitively.

This shift in user expectations raises the question of how to design systems that meet both current and future needs. Involving younger or less experienced pilots early in the process – especially those accustomed to advanced tools – could provide critical insights into aligning systems with the workflows of future operators. Addressing these evolving requirements is essential for ensuring long-term usability and relevance in safety-critical environments.

6. Future Recommendations

The CAIRDA work provided valuable insights specific to the development and validation of AI systems for aviation. Based on these observations, the following recommendations focus on practical considerations for similar future projects:

Engage a Diverse Range of Users To address variability in user preferences and operational contexts, it is important to involve pilots from diverse operational backgrounds (e.g., different aircraft manufacturers and levels of experience). This will help ensure that system designs remain flexible and adaptable to a wide range of workflows and automation philosophies.

Incorporate Feedback Across All Stages of Familiarity Pilots' initial impressions during workshops may not reflect their long-term experience with a system after training and repeated use. Validation efforts should evaluate usability at varying levels of familiarity, balancing early feedback with the potential for pilots to adapt to and appreciate a system over time with proper training.

Anticipate Future User Needs Given the long timelines typical of aerospace projects, it is important to consider how future generations of pilots who may have greater familiarity with advanced technologies, will interact with the system. Involving less experienced or younger users during development may provide insights into how the system will align with future user expectations.

Leverage Alternative Methods When User Access Is Limited Constraints on pilot availability during development can be mitigated by using simulation-based studies, engaging proxy users such

as instructors, or utilizing operational data to inform early design decisions. These approaches can provide meaningful input even when direct user access is limited.

Support Workflow Integration and Evolution AI systems should align with current workflows while also recognizing opportunities for workflows to evolve. Collaboration with operational stakeholders can help ensure that AI is integrated in a way that enhances safety and efficiency, without disrupting established procedures.

7. Conclusion

The CAIRDA work serves as a practical example of how HCD can be applied to develop AI systems for safety-critical environments like aviation. By involving pilots in the design process and focusing on operational needs, the project highlighted how HCD principles can guide the creation of systems that support real-world workflows and address user concerns, such as explainability and usability.

The project also revealed specific challenges including limited access to end users during development, the need to balance first impressions of usability with the impact of training and familiarity, and the importance of anticipating the needs of future users. These findings illustrate how the application of HCD in safety-critical domains requires nuanced approaches to address the unique demands of these environments.

By documenting these experiences and insights, the CAIRDA work contributes to the broader understanding of HCD in practice, offering valuable learnings for future efforts to design AI systems that integrate seamlessly into complex operational contexts.

8. Contact Author Email Address

Amandine Mayima, amandine.mayima@collins.com

9. Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

10. Acknowledgements

JARVIS has received funding from the SESAR Joint Undertaking under the European Union's Horizon Europe research and innovation program under grant agreement No 101114692. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or SESAR 3 Joint Undertaking. Neither the European Union nor SESAR 3 Joint Undertaking can be held responsible for them.

References

[1] C. Van Droogenbroeck, E. Rankova, A. Papenfuss, T. Bos and R. Zon, "Human-AI Teaming – Challenges from a Practitioner's Perspective," in *International Conference on Human-Computer Interaction*, 2025.

[2] EASA, "Artificial intelligence roadmap 2.0," 2023.

[3] C. K. Coursaris, J. Beringer, P.-M. Léger and B. Öz, The design of human-centered artificial intelligence for the workplace, Berlin: Springer, 2025.

[4] J. Boyer, "Human integration design processes (HIDP)," No. S-1169., 2014.

[5] J. C. Mcguire, J. A. Zich, R. T. Goins, J. B. Erickson, J. P. Dwyer, W. J. Cody and W. B. Rouse, "An exploration of function analysis and function allocation in the commercial flight domain," NASA, 1991.

[6] J. M. Canino, L. Gómez, J. García, J. Besada and J. R. Casar, "An agent oriented analysis and modeling of airborne capabilities for trajectory based operations," in *2009 IEEE/AIAA 28th Digital Avionics Systems Conference*, Orlando, FL, USA, 2009.

[7] M. Li, M. Wang, D. Ding and G. Wang, “Development and Evaluation of Single Pilot Operations with the Human-Centered Design Approach,” *Aerospace,* vol. 9, no. 10, p. 601, 2022.

[8] ICAO, “Humans Factor Training Manual,” Doc 9683, 1998.

[9] ICAO, “Manual of Radiotelephony,” Doc 9432, 2007.

[10] ICAO, “Procedures for Air Navigation Services — Air Traffic Management,” Doc 4444, ATM/501, 2001.

[11] FAA, “Airplane Flying Handbook,” FAA-H-8083-3C, 2021.

[12] FAA, “Standard Operating Procedures and Pilot Monitoring Duties for Flight Deck Crewmembers,” AC 120-71B, 2017.

[13] “Checklists - Purpose and Use,” SKYbrary, [Online]. Available: https://www.skybrary.aero/articles/checklists-purpose-and-use. [Accessed April 2023].

[14] “Flight Crew Operating Manual (FCOM),” SKYbrary, [Online]. Available: https://www.skybrary.aero/articles/flight-crew-operating-manual-fcom. [Accessed April 2023].

[15] “Flight Preparation and Conducting Effective Briefings (OGHFA BN),” SKYbrary, [Online]. Available: https://skybrary.aero/articles/flight-preparation-and-conducting-effective-briefings-oghfa-bn. [Accessed May 2023].

[16] M. R. Endsley, “Toward a theory of situation awareness in dynamic systems,” *Human Factors,* vol. 37, no. 1, pp. 32-64, 1995.

[17] N. Brandenburger, A. Dressler and J. Grippenkoven, “DLR Workload Assessment Tool (DLR-WAT) – Official English Version,” 2023.

[18] D. M. Dehn, “Assessing the Impact of Automation on the Air Traffic Controller: the SHAPE Questionnaires,” no. 16(2), pp. 127-146, 2008.

[19] J. Brooke, “SUS—A quick and dirty usability scale,” 1996.

[20] M. Jipp and A. Papenfuss, “Reliability and Validity of the SHAPE Teamwork Questionnaire,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Los Angeles, 2011.