engineering proceedings

MDPI

Proceeding Paper

OLIVIA: Enabling Joint Cognitive Work in Aircraft Divert Scenario Through Operational Intentions †

Ricardo J. N. dos Reis [1,*], Anaisa Villani [1], Silvio Romero Oliveira do Nascimento Filho [2], Charles Dormoy [3], Jaime Diaz-Pineda [4] and Théodore Letouzé [5]

1 Embraer Portugal, 1050-186 Lisboa, Portugal; anaisa.villani@embraer.com.br
2 Embraer SA, São José dos Campos 12227-901, Brazil; silvio.romero@embraer.com.br
3 CATIE (Centre Aquitain des Technologies de l'Information et Électronique), 33400 Talence, France; ca.dormoy@catie.fr
4 Thales AVS, 33700 Mérignac, France; jaime.diazpineda@fr.thalesgroup.com
5 BordeauxINP-ENSC, IMS UMR 5218, Université de Bordeaux, CNRS, 33400 Talence, France; tletouze@ensc.fr
* Correspondence: rjreis@embraer.fr
† Presented at the 15th EASN International Conference, Madrid, Spain, 14–17 October 2025.

Abstract

OLIVIA (**O**perationa**L** **I**ntentions ad**VI**ser for **A**viation) was developed in the HAIKU project. It is a flight deck tool providing support to mission-level decisions in complex situations by assessing and prioritizing route options according to operational intentions. It uses Artificial Intelligence to translate (1) operational intentions from pilots to route generation and optimization inputs and (2) route proposal KPIs into operational intention assessments. This paper reports on the final development of OLIVIA, the results from the human-in-the-loop experiments, and insights and recommendations regarding the development of similar assistants for the flight deck.

Keywords: cockpit assistance; operational intentions; human–AI teaming

1. Introduction

Automation in aviation has led to better operational performance, reduction in workload and a more strategical role for humans, increasingly allowing them to focus more on the "big picture" aspects of safety and efficiency, and less on repetitive, non-rewarding tasks. Those effects are expected to be advanced by the introduction of Artificial Intelligence (AI)-based assistants in flight decks, Air Traffic Controls (ATCs), airline operation centers, and airports. They will support tasks performed by humans, not only providing enhanced information acquisition and analysis capabilities, but also playing collaborative roles—supporting decisions and performing tasks in coordinated settings, enabling them to handle more complex and dynamic scenarios.

1.1. Joint Cognitive Work/Human–AI Teaming

The new/enhanced abilities made possible by AI-based systems—either purely analytical or focused on enhancing human–machine interaction—represent significant changes in joint cognitive systems, i.e., the coupling of humans and machines to perform cognitive work. In this case, the specific form of Human–AI Teaming (HAT) is addressed, namely, adopting the classification proposed in an EASA concept paper [1] (see Table 1). The "teaming" construct is thus adopted to cover specific patterns of joint relationships on task dynamics and the sharing, authority and dependence of joint work between humans

Academic Editors: Spiros Pantelakis, Andreas Strohmayer and Gustavo Alonso

Published: 14 May 2026

 https://doi.org/10.3390/engproc2026133146

and machines and not the personification of the machine. In the EASA concept paper, different levels depend on how authority and task dynamics are allocated between humans and machines.

Table 1. EASA AI levels (adapted from [1]).

L 1: Assistance to Human	L 2: Human–AI Teaming	L 3: Advanced Automation
1A: Human Augmentation	2A: Human and AI-based system cooperation	3A: The AI system makes decisions and performs actions, safeguarded by the human.
1B: Human cognitive assistance in decision and action selection	2B: Human and AI-based system collaboration	3B: The AI system makes non-supervised decisions and non-supervised actions.

Figure 1 maps key research areas for HAT systems according to EASA levels, layering them on top of critical teaming capabilities described in [2]. Although EASA only attributes HAT classification for Level 2, some HAT capabilities start already for Level 1.

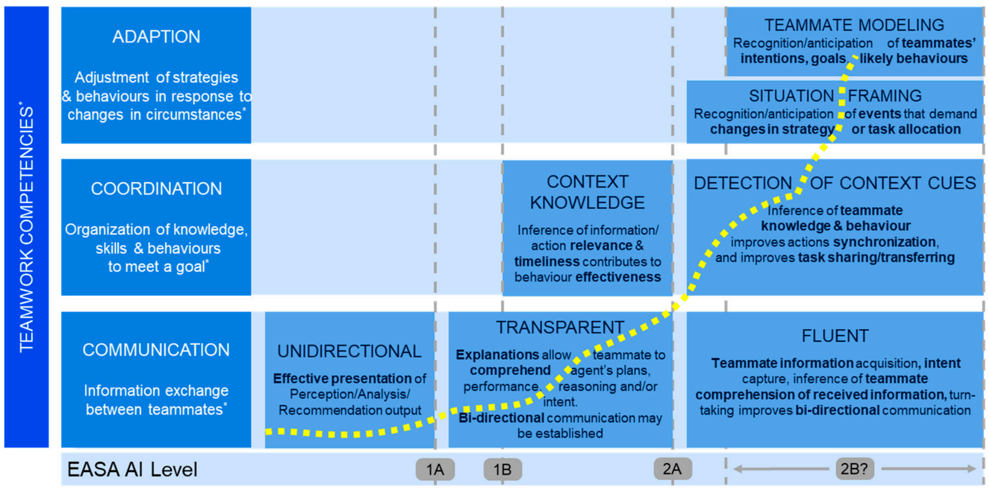

Figure 1. Generic capabilities for teaming, correlated with EASA levels. The yellow line is notional, to represent cumulative teaming capabilities. Definitions signaled with * come from [2].

These capabilities are orchestrated while addressing the following challenges:

- Ensure performance within an operational domain, while addressing issues of learning assurance [1], specifying training, etc.
- Address new forms of work and relationship between humans and machines, enabled by AI technologies.

1.2. What Is HAIKU

HAIKU (2022–2025) [3] was a Horizon Europe project to advance knowledge and tools to develop intelligent assistants in aviation using HAT. Its main goal was to deliver guidelines for the human-centered development of trustworthy AI-based intelligent assistants. Such guidelines resulted from pursuing six different use cases across the aviation domain, sharing human performance, safety, operational and liability assessment methods to support design. The future contexts, challenges and opportunities in each use case were also jointly assessed, including societal acceptance and workforce changes.

1.3. What Is OLIVIA

HAIKU Use Case 2 (UC2) focused in improving the joint cognitive performance for strategical tasks in the flight deck, by enhancing human–machine communication and coordination (Figure 1). It used Thales COMBI technology [4] for human–machine bi-directional communication using operational intentions constructs. This alleviates the users' workload and improves situation awareness in situations where several aspects must be assed to select a course of action. Operational intentions becomes the main communication layer, with COMBI as an intermediator between human high-level goals and machine-specific technical parameters (low-level metrics).

The UC2 assistant was named OLIVIA, for **O**perationa**L** **I**ntentions ad**VI**ser for **A**viation. OLIVIA's role was to provide support for in-flight reroute/divert decision-making in complex scenarios involving weather threats. It resulted from the feedback and insights obtained from activities with pilots, from which OLIVIA's requirements regarding operational and safety objectives were developed. OLIVIA's final concept was classified as an EASA Level 1B assistant, where the teaming capabilities of communication and coordination were exploited. A high-fidelity prototype was developed for the second validation wave (VAL 2), to verify concept compliance with requirements, deploying an initial version of a COMBI translator [5]. OLIVIA was developed by a joint team involving Thales, Embraer, ENSC, and CATIE.

1.4. Driving Questions

This proceeding shares results and insights from the high-fidelity trials of OLIVIA. These were driven by two major research objectives:

OBJ 1. Does OLIVIA enable effective and efficient human–machine communication using operational intentions?

OBJ 2. Do the methodologies developed during the project support safe and effective assurance for HAT assistants?

2. Material and Method

HAIKU used a two-stage validation approach, with incremental realistic human-in-the-loop exercises. In the first phase, the development team focused in interacting with commercial pilots to identify assistance value propositions and develop three assistant concepts with different levels of human–AI partnerships. These were then prototyped as storyboards and tested in low-fidelity simulations with pilots in the first validation wave (VAL 1) [6]. After this, a final concept was defined, prototyped and evaluated through high-fidelity simulations at Thales (FR) facilities in Bordeaux with 10 commercial pilots. A description of the validation 2 stage (VAL 2) material and methodology follow.

2.1. Material

The prototype used in VAL2 was designed to evaluate the compliance of OLIVIA with operational requirements [7] in a subset of operationally relevant scenarios. A COMBI "upward" translation model (producing operational intention assessments based on solutions' performance indicators) was designed using Genetic Fuzzy Trees (GFTs). The synthetic datasets used for supervised learning were generated to be representative of the situations and reroute solutions (trajectories and alternate airports) that could be found on a Marseille–Munich flight, after the destination airport closed due to weather deterioration. Each solution was then annotated using assessments of seven pilots with experience in commercial flights, captured in interviews. A COMBI "downward" translation model (from operational intention prioritization to technical parameters for route generation) was

https://doi.org/10.3390/engproc2026133146

tuned to generate solutions consistently aligned with the declared operational intentions (using the "upward" translator assessments) [5].

The HMI design was driven by human factor considerations, including the operational explainability of AI (OpXAI) attributes and pilot situation awareness (SA) reinforcement. In the current prototype, several explainability layers are already implemented to support pilot understanding of the system logic. These layers indicate how proposed reroute solutions (trajectory + destination) align with operational intentions: (1) a graphical representation of operational intent for each solution, enabling rapid visual assessment; (2) numerical and categorical, three-level color-coded key performance indicators (KPIs), green indicating optimal alignment, yellow denoting acceptable with minor trade-offs, and orange representing acceptable but degraded conditions; those available on demand and providing a transparent rationale for system ranking and trade-offs (red is intentionally excluded, as non-viable options are not presented); and (3) contextual details on candidate destinations (e.g., weather, infrastructure, operational constraints), also available on demand and using the same color-coding logic, allowing pilots to verify and interpret the underlying decision factors consistently.

The prototype was integrated with Thales Flight Management System (FMS) FlytX simulation platform, as illustrated in Figure 2(1). The central display unit integrated FMS and OLIVIA interfaces for flight replanning. OLIVIA provides in its top right frame the current setting of operational intentions, allowing its adjustment (Figure 2(2)). After calculation, operational intentions assessment is given in the top frame, and details (KPIs) of each solution are presented below (Figure 2(3)). The left frame shows an interactive map with information on the current route, proposed solutions, weather and airports. Pilots can also request detailed information for each destination when interacting with this map.

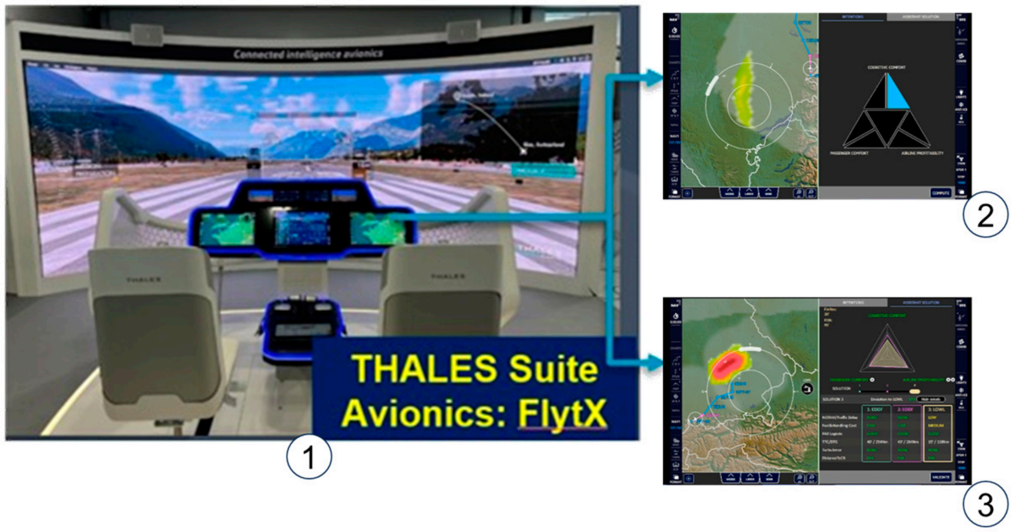

Figure 2. FlytX simulation platform (**1**) and HMI of replanning system integrated with OLIVIA (**2**,**3**).

2.2. Method: Evaluation OLIVIA with High-Fidelity Simulations

In VAL 2, human-in-the-loop simulation with commercial pilots was performed at Thales Innovation facilities in Bordeaux. Variations in an in-flight diversion scenario were evaluated, with and without OLIVIA support. Data was obtained through quantitative measurements (physiological, survey) and qualitative interviews. The scenario was designed to represent a realistic, common, non-emergency operational situation. In such contexts, safety margins are preserved, and crews can balance multiple operational objectives, allowing the study to capture routine multi-criteria decision-making dynamics. While the scenario provides a controlled testbed, the underlying mechanisms are expected to extend to comparable diversion situations. High time-critical emergencies or severely degraded information environments would alter decision priorities and fall outside the present scope. Brief description follows below.

Scenarios:

The scenario is a regional flight from Marseille to Munich (EEDM) on a winter morning, with Frankfurt as the planned alternate. Due to deteriorating weather in EEDM, one extra hour of fuel was added. The simulation begins in-route, two minutes before ATC announces EEDM closure due to a snowstorm and instructs the pilot to hold. Four alternates are proposed: Strasbourg, Linz, Milan, and Frankfurt. The pilot must select one and validate a new flight plan.

Participants:

Ten ATP-license pilots (six women, four men) took part, total flight times from 1400 to 18,000 h (average: 8900 h). Eight had captain experience, average age was 40.5 y.

Protocol:

Each session lasted 3 h 20. After a briefing, two practice runs (with and without OLIVIA), and a baseline were performed, with participants using eye tracker. They then completed four 20 min scenarios, two with and two without OLIVIA. Weather and airport indicators varied across runs. With OLIVIA, pilots selected their intentions and received three suitable route proposals. Without OLIVIA, they manually plotted their diversion route on the map interface.

Questionnaires and interviews:

At the end of each run, three task-related questionnaires were administered: workload (NASA-TLX; [8]), situational awareness (SART; [9]), and recommendation transparency [10]. At the end of the experiment, four additional questionnaires assessed trust [11,12], usability (CSUQ; [13]), and acceptability (AIDUA; [14]). Finally, the pilots took part in a self-confrontation interview supported by video recordings of the four runs, in which they were encouraged to reflect on their experience and verbalize the reasoning behind their actions.

3. Results

As mentioned, both quantitative and qualitative results were obtained. A brief overview of the main insights from quantitative results is provided. Qualitative results are further detailed in the interview analysis section.

Highlights from quantitative results:

The results of the questionnaires allow us to confirm that the assistance provided by OLIVIA did not generate any additional mental load, that awareness of the situation remained sufficient to complete the task, and no issues of trust, usability or transparency were identified. The quantitative results from the questionnaires are presented in a separate manuscript, which is currently in the process of publication. This additional work provides a detailed statistical analysis that complements the findings discussed here.

Highlights from interviews:

The main insights that resulted from pilots' feedback are briefly described in Table 2, grouped by theme (with some overlaps between them). These were captured as affecting (or being affected) by the affordances created by OLIVIA HMI design language.

Table 2. Main topics and insights captured from the thematic analysis of the interviews.

Theme	Topics and Insights
Human–AI collaboration in decision-making processes	• Different decision-making styles • Different solutions exploration preferences • Considerations on situation variability (e.g., time criticality) • Relevance of offering diverse ways to interact with the system
System usability and interface design	• Color code standardization improves information processing • Lack of clarity in color code meaning for KPIs • Preferences for ranked solutions in a list • Graphical representation intuitiveness of operational intentions should be improved
Trust in the AI-based assistance	• Relevance of training and understanding to establish trust • Pilots perceive OLIVIA as an augmentation tool, not as a replacement for human judgement, i.e., OLIVIA solutions must be complemented by pilot expertise judgement before final decision.
Explainability and transparency	• Relevance of understanding system inner logics, including which information is used and how it is processed, to properly use the system • Interest in obtaining reasoning behind solutions
System limitations and areas for improvement	• Concerns with incomplete information provided on weather/KPIs. • Concerns on weather analyses limitations, mostly in information validity and dynamics prediction
Training and adoption considerations	• Insufficiency of training provided in the exercise to build familiarity, understanding of system capabilities and trust • Value of incorporating OLIVIA in training to develop expertise • Extra concerns in providing adequate training for new pilots

Table 2. *Cont.*

Theme	Topics and Insights
Added value and operational impact	• Reduction in complexity to find out operational diversion solutions • Improved SA and decision-making enabled by gathering and integration of more comprehensive information • Improved workflow in diversion scenarios • Potential to reduce workload in high-pressure situations
Suggestions and recommendations	• Provide integrated wind information • Provide details on/justifications for KPIs • Provide ranked lists of available airports, with more options • Avoid presenting multiple solutions for the same airport • Enable pilot to submit own solution for system assessment • Clarify (in training) how the system ensures flight safety

Key insights about human–AI interaction in this type of decision support system include: the conditionality of trust in understanding system logic and performance consistency; the perception of OpXAI as a crucial element for the appropriate use of the system; necessary trade-offs between information comprehensiveness and conciseness, to avoid overwhelming; the expectations that assistants like OLIVIA will enhance pilots' SA and decision quality, when integrated with their own expertise, which will keep its current relevance and cannot be replaced.

4. Final Considerations

Final considerations regarding OBJ 1 (*"does OLIVIA enable effective and efficient human–machine communication through operational intentions?"*) and OBJ 2 (*"did the methodologies used in UC2 provide assurance for the development of HAT assistants?"*) follow.

4.1. OLIVIA Final Considerations (OBJ1)

Overall, VAL2 showed the **OLIVIA** concept's **ability to support crew decision-making** through the **easy integration of relevant information**, **fluid communication**, and **efficient navigation between different abstraction levels**. Pilots liked the **system's agility**, **ability to quickly provide solutions** (i.e., within seconds), as well as its **ergonomics**, which they perceived favorably and as being conducive to operational adoption. These results confirm the relevance of assistance focused on user priorities and a clear structure of the situation.

Improvement opportunities remain as: **latent explainability** (created through training and familiarity); **visual clarity** (interface elements constructs); **diversification of data**. The integration of more intuitive representations, additional explanatory layers, and customization options is now a key step in enhancing the transparency and effectiveness of assistance. Qualitative pilot feedback indicated the benefit, particularly in a training context, of better understanding how specific KPIs are constructed: they expressed interest in clearer visibility into the operational factors contributing to these indicators (e.g., weather conditions, accommodation availability, or operational constraints), as well as explicit clarification of the meaning associated with each color-coded KPI.

Overall, OLIVIA appears to be a promising concept for decision support in the aeronautical context, providing a solid basis for future developments towards more robust and intelligible human–AI collaboration.

4.2. General Considerations for HAT Development (OBJ2)

Although not explored in detail in this proceeding, some thoughts are offered regarding the development methodology, following work reported in [5]. Human-centered design can be a truism when the final goal is safety, efficiency and value creation. The HAIKU iterative approach with end-users has guided OLIVIA since its inception. This, in the authors view, is critical. The COMBI translator's effectiveness was tightly dependent on the user training, and VAL1 and VAL2 evaluations, which can be taken as a strong assumption for other HAT components. These results are coherent with the other HAIKU use cases [15].

To deliver HAT expected value requires the implementation of specific enablers, as robust human–AI communication mechanisms, and situational awareness joint coherence (i.e., *"shared situational awareness"*), providing a true "cognitive middleware" for cockpit assistance. Human–machine modeling should be further advanced, to accelerate space design exploration with users in the iterative process. Without an efficient HAT engineering process—providing performance assurance on the overall HAT—it is challenging to usher HAT innovation into operations, either in the flight deck or other safety-critical domains. Together, these elements open the way for HAT solutions in aviation.

Author Contributions: Conceptualization, R.J.N.d.R., A.V., S.R.O.d.N.F., T.L., C.D. and J.D.-P.; methodology, C.D. and T.L.; software, J.D.-P., C.D., S.R.O.d.N.F. and T.L.; validation, C.D., T.L., A.V. and S.R.O.d.N.F.; formal analysis, S.R.O.d.N.F. and C.D.; investigation, all authors; resources, J.D.-P.; data curation, C.D. and S.R.O.d.N.F.; writing—draft preparation, R.J.N.d.R., C.D., A.V. and T.L.; writing—review and editing, J.D.-P. and R.J.N.d.R.; visualization, T.L. and C.D.; supervision, J.D.-P. and R.J.N.d.R.; project administration, C.D. and T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding by the European Union's Horizon Europe research and innovation program under Grant Agreement no. 101075332.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: All public data is available in project page in the CORDIS database: https://doi.org/10.3030/101075332.

Conflicts of Interest: Author Ricardo J. N. dos Reis and Anaisa Villani were employed by Embraer Portugal. Silvio Romero was employed by Embraer S.A. Jaime Diaz-Pineda was employee of Thales AVS (FR). The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ATC	Air Traffic Control
EASA	European Aviation Safety Agency

HAT	Human–AI Teaming
KPI	Key performance indicator
OpXAI	Operational explainability of AI
SA	Situation awareness
VAL 1	First wave of validations in HAIKU project
VAL 2	Second wave of validations in HAIKU project

References

1. EASA. *EASA Concept Paper: Guidance for Level 1 & 2 Machine Learning Applications, Issue 2*; EASA: Cologne, Germany, 2024.
2. Stowers, K.; Brady, L.L.; MacLellan, C.; Wohleber, R.; Salas, E. Improving Teamwork Competencies in Human-Machine Teams: Perspectives from Team Science. *Front. Psychol.* **2021**, *12*, 590290. [CrossRef] [PubMed]
3. HAIKU Fact Sheet. 2025. Available online: https://cordis.europa.eu/project/id/101075332 (accessed on 10 November 2025).
4. Hourlier, S.; Diaz-Pineda, J.; Gatti, M.; Thiriet, A.; Hauret, D. Enhanced dialog for Human-Autonomy Teaming-A breakthrough approach. In Proceedings of the IEEE/AIAA 41st Digital Avionics Systems Conference (DASC), Portsmouth, NH, USA, 18–22 September 2022. [CrossRef]
5. Minaskan, N. HAIKU Deliverable D4.3: AI Model Performance Report. HAIKU Project. 2024. Available online: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e51159a7cd&appId=PPGMS (accessed on 10 November 2025).
6. Hentati, T.; Letouze, T.; Dormoy, C.; Diaz-Pineda, J.; Reis, R.; Villani, A.; Andre, J. Optimizing Decision Making in Aviation: A New Communication Paradigm for Rerouting. In Proceedings of the 2nd International Conference on Cognitive Aircraft Systems (ICCAS), Toulouse, France, 16–17 May 2024. [CrossRef]
7. Westin, C. HAIKU Deliverable D6.3: Updated Validation Strategy Plan. HAIKU Project. 2024. Available online: https://cordis.europa.eu/project/id/101075332/results (accessed on 10 November 2025).
8. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Adv. Psychol.* **1988**, *52*, 139–183. [CrossRef]
9. Taylor, R.M. Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations, Seuilly-sur-Seine, France, 2–6 October 1989.
10. Hellmann, M.; Bocanegra, D.C.H.; Ziegler, J. Development of an Instrument for Measuring Users' Perception of Transparency in Recommender Systems. In Proceedings of the ACM IUI WORKSHOPS 2022, Helsinki, Finland, 21–22 March 2022. [CrossRef]
11. Ashoori, M.; Weisz, J.D. In AI We Trust? Factors That Influence Trustworthiness of AI-infused Decision-Making Processes. *arXiv* **2019**, arXiv:1912.02675.
12. Jessup, S.A.; Schneider, T.R.; Alarcon, G.M.; Ryan, T.J.; Capiola, A. The Measurement of the Propensity to Trust Automation. In Proceedings of the 11th International Conference on Virtual, Augmented and Mixed Reality, VAMR 2019, Held as Part of the 21st International Conference on Human-Computer Interaction, HCII 2019, Orlando, FL, USA, 26–31 July 2019. [CrossRef]
13. Lewis, J.R. Measuring perceived usability: The CSUQ, SUS, and UMUX. *Int. J. Hum.–Comput. Interact.* **2018**, *34*, 1148–1156. [CrossRef]
14. Gursoy, D.; Chi, O.H.; Lu, L.; Nunkoo, R. Consumers acceptance of artificially intelligent (AI) device use in service delivery. *Int. J. Inf. Manag.* **2019**, *49*, 157–169. [CrossRef]
15. HAIKU. HAIKU–AI in Aviation Highlights, the HAIKU Viewpoint, HAIKU. 2025. Available online: https://haikuproject.eu/haiku-final-glossy-report/ (accessed on 12 September 2025).