Human System Exploration for the AI-based flight deck decision support system IPAS

Jakob Würfel[1] & Frank Flemisch[2]
[1]German Aerospace Center (DLR), Brunswick, Germany
[2]RWTH Aachen University, Aachen, Germany

Abstract

The integration of human users with Artificial Intelligence (AI)-based systems is crucial for the development of AI assistance systems in safety-critical work environments. The German Aerospace Centre (DLR) is developing the Intelligent Pilot Advisory System (IPAS) to demonstrate and research the application of AI in flight decks. The development of the IPAS started with a number of high-level requirements and general ideas for use cases. In parallel with the system development, research focused on how to provide AI explanations tailored to the end user, in this case airline pilots. In order to make creative and effective use of the freedom given and to identify the requirements for explainable AI in the flight deck, the IPAS was developed and researched in an explorative manner following the Human System Exploration (HSE) paradigm from initial vision to an executable demonstrator in a simulation environment. This paper outlines the reason for choosing the HSE paradigm, the modifications made and its practical application. It describes the development and research process, introduces the methods used, presents the results and discusses the value of this approach in terms of scientific research and the development of the IPAS.

Introduction

Recent advances in Artificial Intelligence (AI) technologies have resulted in AI applications not only becoming more present in our everyday lives, but also increasingly used for complex activities and tasks, such as automated driving (Masley et al., 2023; Perri, 2023). In aviation, there is great potential for new AI systems, which are being discussed for example for the flight deck, to support certain tasks, improve work performance and increase safety (Miller et al., 2023). At the German Aerospace Centre (DLR), the LOKI[1] project emphasises a human-centred approach to understand the interaction and collaboration between humans and AI-based systems, to derive guidelines and to develop systems specifically for air traffic controllers and airline flight decks (Stefani et al., 2023). The integration of AI into socio-technical systems like the flight deck is a major challenge, as each group of end-users has different requirements. The concepts of Human Systems Integration (Booher, 2003), balanced Human Systems Integration (e.g. Flemisch et al., 2013) and

---

[1] A preliminary introduction to the LOKI project can be found under the URL http://www.loki.dlr.de

In D. de Waard, D. Manzey, K. Brookhuis, F. Siebert, K. Karrer-Gauß, S. Winkler, F. Di Nocera, A. Toffetti, and T. Franke (2024). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2024 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Human-AI Teaming (National Academies of Sciences, Engineering, & Medicine, 2022) focuses on systems in which the AI system is perceived as a collaborative team member, with an emphasis on communication and coordination to achieve better team performance. System features essential for human-AI systems include team situation awareness, explainability of AI, and calibrated trust. (Endsley, 2023; Ribeiro et al., 2016; Lee & See, 2004)

The integration of novel technologies, such as AI, into complex work environments sometimes requires researchers to develop these system prototypes, mock-ups or demonstrators. This is especially true in early stages, when the technology is not yet fully developed, but research on the integration and impact of these technologies in socio-technical systems should already be performed. In this way, the use of these new technologies can be demonstrated and tested in a realistic setting, providing information on the impact of the technology on human system integration and human factors aspects. (Flemisch et al., 2013; Stanton et al., 2017; Doshi-Velez & Kim, 2017) An example of this is the Wizard of Oz method, where the output of a supposed AI system is generated by a human in a hidden way. (Dahlbäck et al., 1993) When it comes to the integration of AI, especially in critical work environments such as a flight deck, human factors research is important to support safety and usability (Kirwan et al., 2020).

Known processes that structure the development of such systems are used to help developers create these systems. For example, user-centred development models (e.g. ISO 2019) are used to integrate humans and machines to evaluate new systems in the context of human factors. The issue of how to develop new systems for research purposes is addressed, for example, by Parnell et al. (2021) or Plant et al. (2023) in their research on future cockpit designs and technologies. Parnell presents a user-centred development model for researching and evaluating new technologies in the aircraft cockpit. She outlines the steps involved, the stakeholders - who should be involved in development and evaluation - and possible methods.

To research and demonstrate the integration of AI into the flight deck the Intelligent Pilot Advisory System (IPAS) has been under development at DLR since 2021. (Djartov et al., 2022; Würfel et al., 2023; Würfel et al., 2024) At the outset, there were numerous general visions and ideas for the system and for possible research topics. This was beneficial as it allowed the IPAS to be developed from scratch, with plenty of room for innovation and creativity. The initial focus of the research emphasised the challenge of how to achieve explainability of the system's AI output tailored to the needs of airline pilots in a specific use case. An innovative and explorative development process was needed to take the first steps on the unfamiliar terrain, to exploit the potential of the freedom given, to explore how to achieve explainability of AI, and to combine the development of the system itself with the research. This process should guide the development of the new system, contribute to answering the research question, involve the human user and effectively deliver results in the time available, as the research question should be answered in the context of a doctoral thesis. Human System Exploration (HSE) (Flemisch et al., 2022) was chosen as a development paradigm that largely met these requirements for exploration and development.

This paper outlines the process of selecting, adapting and executing the IPAS research and development process, from the initial sketch to the first running demonstrator in the simulator. Finally, the paper concludes on the usefulness of the explorative approach to system development and addressing research.

The Intelligent Pilot Advisory System (IPAS) and its AI-Crew Interaction System (AICIS)

The IPAS is a research project at DLR that aims to explore, research and demonstrate the integration of AI into the aircraft cockpit and is divided into the AI-Crew Interaction System (AICIS), which covers the human-system integration aspects, and the AI Core Module (AICOM), which contains the AI of the system. When the development of the IPAS began, there were general objectives and use cases for the system. The high-level requirements for the IPAS were that it should integrate AI technologies and provide advice and decision support in emergencies adapted to the current situation.

In its AI Roadmap 2.0 (EASA, 2023), the European Aviation Safety Agency (EASA) calls for a human-centred approach for the application of AI in aviation. In the human factors section of the roadmap, there is a requirement for "operational explainability of AI", which means that AI systems should be able to provide the human end-user with understandable information about how the AI generated its results. These explanations should be tailored to the needs, expertise, tasks and goals of the end user. Better explainability is a system quality that enables the end user to effectively use and manage the system (Gunning et al., 2021), but also, for example, enables the user to achieve calibrated trust in the system (Endsley 2023). Following the EASA requirement, the AICIS research focuses on how to achieve AI explainability tailored to the needs, expertise, tasks and goals of airline pilots through a user-centred design approach.

The Human System Exploration Paradigm

Exploration, the key concept of HSE, means investigating, studying, analysing and trying out new ideas and scenarios, similar to exploring new worlds. (Flemisch et al., 2019) Following an explorative approach, developers do not know all the parameters at the start of development and are open to innovation and unexpected discoveries, learning by trying new ideas, taking wrong turns and trying new approaches. In the end, the probably best system design or prototype can be found. The Human System Exploration (HSE) paradigm is visualised by the process shown in Figure 1, known as Innovation Turbine. It uses the form of an exhaust jet of a supersonic engine (“Mach turbine”) as a metaphor. The exploration takes place in the area of the expanding Mach diamonds, while the definition and focussing on a solution takes place in the merging area of the Mach diamond. In the Mach disc area, a kind of synthesis of the explored results takes place before a new exploration begins. The innovation turbine consists of three major phases: ideation, in which visions and ideas are collected and specified, system analysis, in which the context of the system is analysed and, for example, use cases are defined and a system model is developed, and thirdly the design and development phase, in which prototypes are developed and tested. The model is described in detail in (Flemisch et al., 2022). The model’s iterative and participatory design approach is an important feature, as it allows for early and iterative exploration of new human-machine systems with users and stakeholders. This allows early development of systems, a focus on balancing different requirements, ideas and methods, and the promotion of creativity and innovation. The application of the turbine is illustrated by the example of human-automated vehicle interaction in (Usai et al., 2023).

Setting up a development process for the IPAS

In the initial phase of IPAS development, the selection of a process model was crucial. This model had to balance the tandem of research and development, leaving the necessary room for creative exploration, but effectively approaching results within the timeframe of a doctoral thesis. The IPAS should be both a completely new system, based on a collection of visions and ideas, demonstrating operational AI in the flight deck, and a research platform to explore and answer several research questions. The first focus was on researching the operational explainability of AI, tailored to the needs, expertise, tasks and goals of the future end users of the IPAS, the airline pilots, and finally how these findings could be implemented and validated using an AICIS demonstrator. At the start of development, there were only few principles, examples or even guidelines for applying or explaining AI on the flight deck, so these had to be explored. In line with the EASA roadmap, a user-centred research approach was required. The team therefore needed a user-centred process that would allow creativity and innovation to transform conceptual ideas and visions into a tangible and functional system, while exploring approaches to achieve AI explainability, taking into account existing scientific knowledge.

Choosing and adapting the HSE process for the IPAS development.

The aforementioned prerequisites for the development of the IPAS resulted in the selection of the HSE paradigm and associated process model because of its explorative and user-centred approach. In addition to developing the system, the explorative approach also promised insights into the research question of how to make AI explainable to pilots. In order to achieve a structured and organised development process despite the high degree of freedom and exploration, and to select methods for each phase that have already proven useful in the development of new flight deck systems, the HSE Innovation Turbine was combined with a framework specifically for user-centred design of flight deck technologies presented in (Banks et al., 2018) and (Parnell et al., 2021). Figure 1 shows, the combination of the HSE model and the framework.

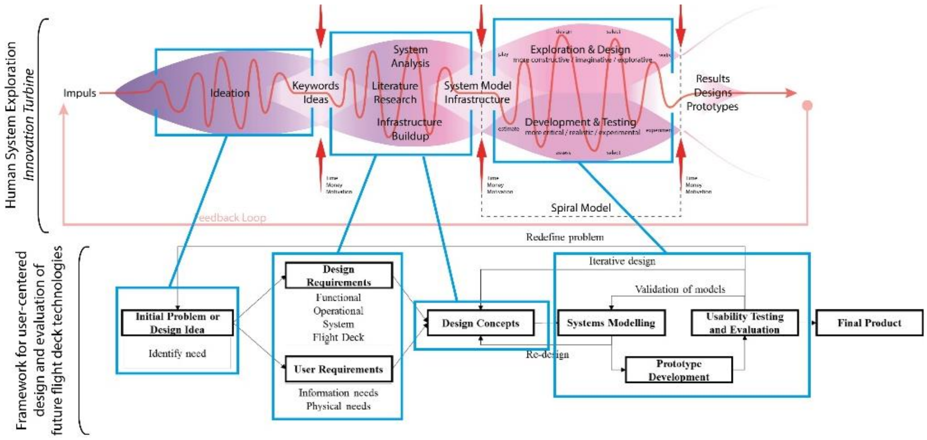

Figure 1. The graphic above shows the HSE model, the turbine shape is recognizable, below the framework for the user-centred design of new flight deck technologies. The authors have marked the areas of the two approaches that will be combined for the development of the IPAS.

The sections of the framework could be mapped to the three phases of the Innovation Turbine to add structure by defining the phases in more detail, specifying work packages and their deliverables. The explorative HSE approach is still maintained through many iterations, loops and room for creativity and innovation. Finally, the development process shown in Figure 2 was defined as a synthesis of HSE and the User Centre Flight Deck Technologies model. The explorative phases, the Mach diamonds are retained, while the turbine has been modified with the design flow and milestones from the framework for the user-centred development of flight deck technologies. The IPAS had to be defined and developed at a general level before the subsystem AICIS could be developed in more detail. Therefore, the development should become more detailed with each iteration and loop, delving deeper into the research question.

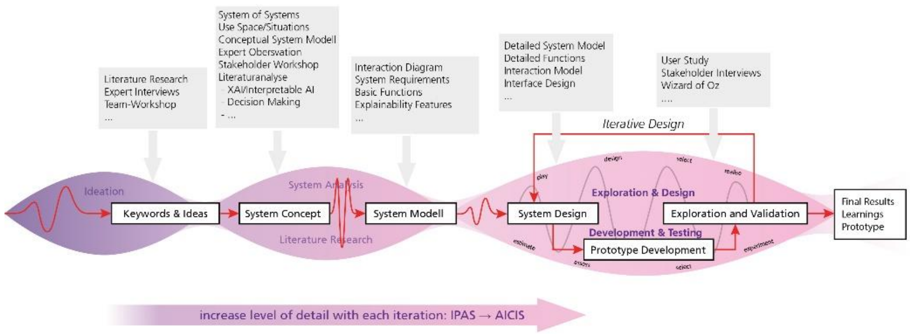

Figure 2. The process model defined for the development of IPAS and the AICIS, which is a combination of the HSE innovation turbine and a framework for the user-centred design of future flight deck technologies.

Methods and Results

In the following section, the development phases of the process model and the methods applied are described on the basis of specific examples.

Phase 1 - Ideation

The ideation phase was driven by the vision of developing an AI-based assistance system to support pilots during emergencies, with a particular focus on ensuring the explainability of AI features. This phase was primarily characterised by literature research, which focused on exploring current research topics related to explainable AI. In addition, initial expert interviews were conducted to assess the existing needs for support on the flight deck. These activities laid the groundwork for the development of use cases and system functionalities.

During this phase, two main applications for AI-based assistance systems in the flight deck were identified: 1. Support for strategic in-flight planning during normal operations. 2. Support for decision-making during emergencies, particularly in identifying and evaluating alternative airports (Würfel et al., 2023).

The expert interviews also revealed pilot concerns about the use of AI, as well as general user requirements for AI systems. For instance, pilots expressed concern that they might struggle to understand the AI-generated results or that interpreting these results could demand too much cognitive effort (Würfel et al., 2023). Additionally, both the interviews and literature review confirmed that research into explainable AI systems, specifically tailored to end-user needs, was essential and needed further exploration. This supported the initial focus and highlighted the need to prioritise explainability in the IPAS development. The outcomes of the literature review and expert interviews were discussed during a team workshop, leading to the formulation of ‘Keywords and Ideas’. This document marked the transition to the next phase and served as the foundation for IPAS development. It outlined the basic system concept, top-level use cases, and preliminary requirements for the system (Würfel et al., 2023):

The IPAS should monitor the status of the mission at a strategic level, highlight anomalies, and provide assistance when necessary.

- IPAS should function as a decision-support system, with recommendations generated by AI algorithms.
- The system should offer particular support during emergencies or abnormal procedures, specifically in the FOR (Facts, Options, Risks) part of the FORDEC decision-making model, to assist pilots in high-workload situations.
- AI algorithms should be employed to identify and evaluate operational options, such as alternate airport choices, depending on the situation.
- AI-generated results should be presented in a way that allows pilots to easily interpret them, fostering trust in the IPAS system, and enabling pilots to understand the recommendations both effectively (enhancing performance) and efficiently (in a timely manner), particularly in high-stress situations.

Phase 2.1 - System Analysis and System Model for the high-level system IPAS

The main objective of the second phase was to conduct a systems analysis of the socio-technical system, which would form the basis for both research and system development. The approach was to analyse and dissect the reality in which the system will operate in order to create a system model and later, in Phase 3, to integrate and test the new system. Knowledge for the analysis was gathered through individual interviews with experts, including airline pilots, aeronautical engineers, and psychologists, alongside an extensive literature review focused on aeronautical decision-making procedures, flight deck systems, explainable AI, and user interfaces. Based on the insights gained from these interviews and the literature, a "system of systems" was developed to provide a high-level overview of all subsystems and stakeholders involved in in-flight decision-making situations where the IPAS would be implemented. This system of systems not only outlined the relationships between these systems and stakeholders but also highlighted the placement of the IPAS within this complex context.

Detailed use cases were created to describe the specific tasks and functions the system should perform, detailing the actions required from the crew, as well as identifying triggers and outcomes. For instance, one use case, illustrated in Figure 3, describes a scenario where the crew must decide to divert to an alternate airport due to a fire in the cargo compartment, assuming there are no technical limitations.

Use Case Number	UC-DecisionSupport-1		
Use Case Name	Decision Support: Land ASAP		
Brief description	IPAS automatically detects that the aircraft should land as soon as possible and provides suitable alternate airport options. The priority of the evaluation of suitable airfields is the flight time needed to reach the options.		
Example Scenario	Fire in cargo compartment		
Source	Internal Report "IPAS-PilotWorkshop2021": 3.2 Use Case #1; 3.3.2 Statement #14		
Precondition	Aircraft is in cruise; FORDEC for ABNORM or Emergency Procedure; No operational limitations due to technical condition		
Trigger	An event occurs that forces the crew to land as soon as possible		
Ordinary Sequence	**Step**	**Action Crew**	**Action IPAS**
	1	-	Situation detected by IPAS, need and type of support identified (Land ASAP)
	2	-	Identify and assess different alternative options with regard to the detected situation.
	3	-	Provide alternate airport options to land as soon as possible at a suitable airport.
	4	Check and Discuss and probably follow IPAS recommendations	-
	5	Inform the IPAS about decision	-
	6	-	Provide information on the selected option.
	7	Check und discuss IPAS information, draw a conclusion	-
Postcondition	FORD is completed, the crew executes the selected decision.		

Figure 3. Example Use Case of the IPAS. First published in Wuerfel et al. (2023). Template according to Santos et al. (2015), slightly adapted.

The initial sketch of the system model began with simple drawings to illustrate how the system functions, building upon the identified use cases and the system of systems. This first sketch, shown in Figure 4, detailed the order and type of functions to be realised and outlined how pilots could interact with the system. This sketch provided developers with a shared understanding of the system and served as a foundation for more detailed development in subsequent iterations.

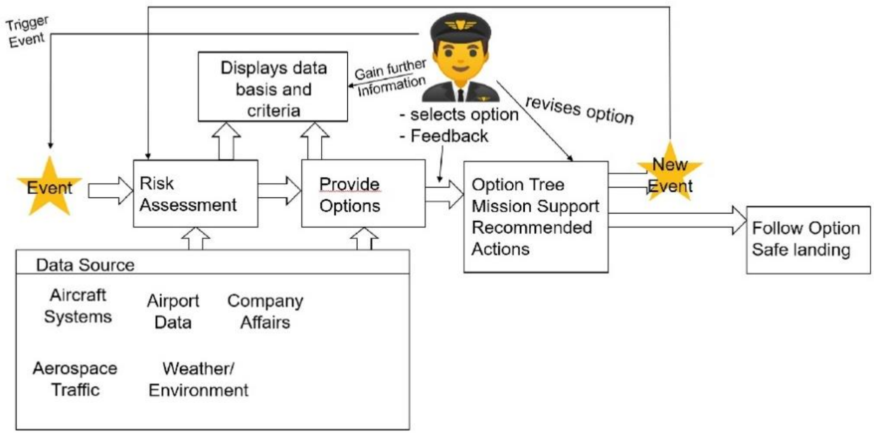

Figure 4. The figure shows the first initial system concept of the IPAS, created a few weeks into development. The sketch shows the first ideas of how the system could operate and how the interaction between the pilot and the system could work.

Several alternative system concepts and sketches were developed throughout the phase, starting from this first system concept. These concepts refer to a basic and conceptual description of the system or individual aspects of its functionalities. Although the initial models were simple and schematic, they provided a valuable basis for later system functionality and visualisation to further develop these early concepts into a refined system model, a set of basic functions was identified and assigned to the subsystems, ensuring they could fulfil the defined use cases. These functions were iteratively combined with findings from the analysis and literature review, leading to the creation of a final system model, as depicted in Figure 5. This model provides an abstract representation of the IPAS and its subsystems, AICIS and AICOM, facilitating a higher-level understanding of the system without focusing on specific details. The level of automation was also defined according to the framework established by Parasuraman (Parasuraman et al., 2000; Save et al., 2012) to complete the description of the system.

To validate the basic functions of the IPAS and gain initial insights into the explainability needs of airline pilots, a user study was conducted using a rudimentary click-dummy. The Wizard of Oz method (Dahlbäck et al., 1993) was employed to simulate AI-based results, allowing participants to engage with the click-dummy while thinking aloud, thus providing insights into their mental models (Hoffmann et al., 2023). Based on the results, the system of systems, the basic functions, the keywords and ideas, and the system model were extended and reworked.

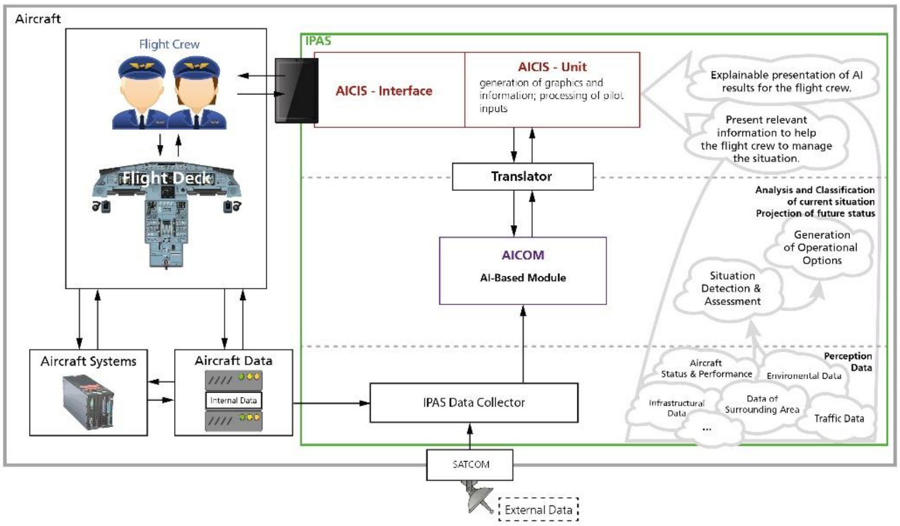

Figure 5. The system model of the IPAS, which shows the interaction between the crew and the subsystems, and also shows which process steps are performed in each part of the system. This system model can be traced back to an initial sketch shown in Figure 5. The model was first presented in Würfel et al. (2023).

A key outcome of this phase was the development of a concept for achieving operational explainability for AI-generated outputs. High-level requirements for the Explainable User Interface of the AICIS were derived from the findings of a study published in Würfel et al. (2024). According to this concept, AI results should be made explainable through individual chunks of information—excerpts from AI-based calculations that pilots can interpret. This approach allows pilots to understand the rationale behind the generated results through self-explanation, using their own knowledge in combination with the provided information, without needing to fully understand the underlying algorithm. This approach aims to balance decision performance with adequate explainability. A principle sketch of the approach is shown in Figure 6.

Phase 2.2 - System Analysis and System Model for the subsystem AICIS

Once a system model for the overall IPAS and functional descriptions for the AICIS and AICOM subsystems had been developed, indicating a sufficient understanding of the IPAS, the system analysis phase was restarted with a specific focus on the AICIS subsystem. This analysis focused on understanding the pilots' decision-making processes, explainability and information needs, using the results of studies and interviews conducted and a review of relevant literature. The primary objective of this phase was to model the AICIS, define system requirements, and develop features that enhance the explainability of the AI, in line with the operational explainability concept defined in Phase 2.1.

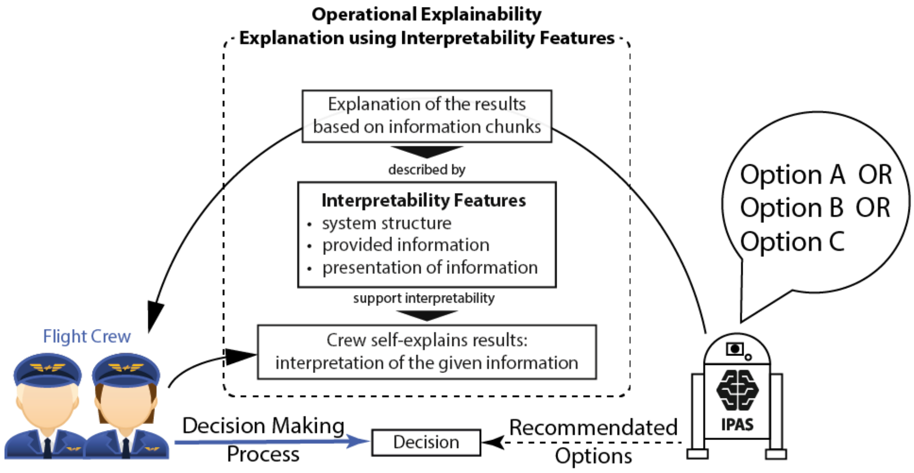

Figure 6. The principle of how explainability is to be achieved can be seen. The explainable interface provides information chunks from AI-based results. How this information should be presented and what content they have is defined by the interpretability features. With the help of these information, the crew can explain the results to themselves by interpreting the information and combining it with their own knowledge. According to Würfel et al. (2024)

During this phase, detailed system tasks were identified, outlining the functions that the AICIS should perform. This description was iteratively refined throughout the analysis. Several specific explainability features were developed to implement the defined concept of operational explainability. Additionally, an analysis of the pilots' decision-making process was conducted to ensure that the workflow and interactions between the human operator, the system, and its subsystems aligned with established decision-making models. The outcome of this analysis is presented in Figure 7.

The methods employed included Rasmussen's ladder of decision-making (Rasmussen, 1974; Banks et al., 2020) and Neisser's perceptual circle (Banks et al., 2021). Ultimately, building on the high-level IPAS system model, the workflow diagram, and the basic functions assigned to the AICIS, a detailed system model for the AICIS was created. Implementation requirements were defined, which, when combined with the explainability features, served as the foundation for the development of the demonstrator and marked the transition to the next phase of the innovation turbine.

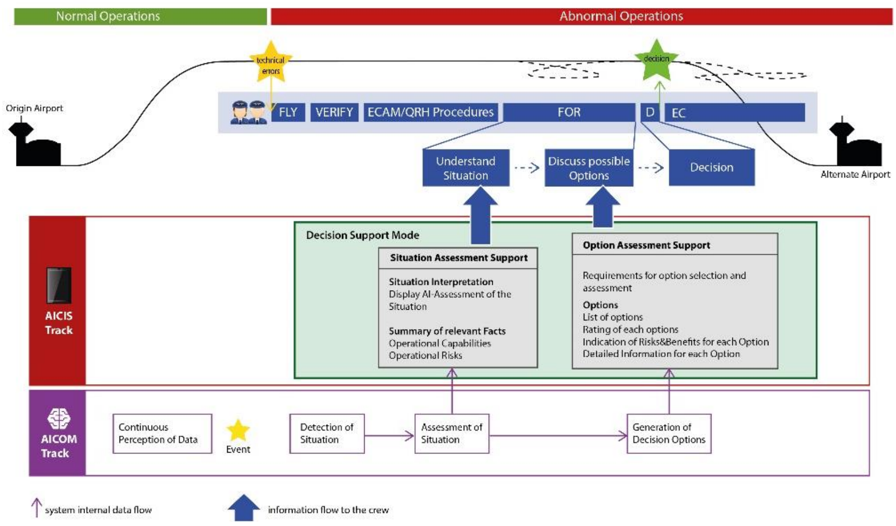

Figure 7. This diagram shows the work performed in the cockpit in an emergency, the functions provided by the AICIS and the AICOM, the workflow and the interaction between the parts of the system. This diagram is also derived from an early system concept shown in Figure 4.

Phase 3 - Explorative Design, Development and Testing

The final phase focuses on the iterative design and implementation of the demonstrator, followed by testing with end users. This process aims to provide both scientific insights and a functional demonstrator. The simulation environment for the IPAS demonstrator is illustrated in Figure 8.

Initially, a swim lane diagram was created to detail and assign functionalities to individual pages of the interface. The implementation process began with initial sketches and progressed through numerous iterations to a final design that integrated the interface, system features, and design considerations derived from the use cases, requirements, and the previously defined explainability features. This iterative approach, guided by workshops, expert reviews, and the development and selection of various design options, culminated in the final implementation of the demonstrator within the flight deck of an Airbus A321 simulator (see Figure 9).

Subsequently, the demonstrator was tested in a study involving seven airline pilots. The aim of the study was to verify the use case, identify usability issues and potential additional functions, validate the existing system design, and pay particular attention to the explainability of the AI-generated results. Methodologically, the experiment began with an exploratory phase in which participants were asked about their ideas for AI-based flight deck assistance, their hopes and concerns regarding AI, and their visions of what such a tool could look like. Following this, the demonstrator was tested by the pilots in five different emergency scenarios in the simulator. Finally, the study was completed with questionnaires and a semi-structured interview.

Usability was assessed using the Post-Study System Usability Questionnaire (Lewis & Sauro, 2021; Stanton et al., 2017), along with several custom questionnaires targeting specific system and explainability features. The semi-structured interviews explored usability issues, the individual need of the system itself, the validity of the aeronautical system's output and explainability features. These interviews were transcribed and analysed using an affinity diagram (Lucero, 2015).

Figure 8. The simulation environment at DLR in Brunswick houses the iSim, a flat-panel simulator that works with touch screens. This allows rapid prototyping without the need to modify hardware.

The analysis of the study focused on three key areas: 1. usability and feedback on the interface, tools and features; 2. comments on the quality of the AI-based airport ratings; and 3. feedback on explainability features designed to help users understand the AI-generated results. Participants agreed that the tool was helpful in selecting a suitable airport and positively influenced their decision making by providing more options and considering additional factors. However, some suggestions were made for interface improvements and additional features. For example, feedback on the interface highlighted areas such as revising the colour codes for the ratings, making the wording more consistent with pilot terminology, and making the map display optional to save space. Suggestions for additional features focused on improving the tool's ability to predict conditions at the destination, in particular by providing an indication of available 'Plan B' options and estimating fuel reserves when approaching the selected alternate airport. During the interviews, participants discussed the explainability features implemented in the AICIS and identified potential improvements, but generally expressed that they were able to understand the information provided by the AI. One example of a misleading explainability feature was the use of small green and red arrows to indicate whether a factor was important or less important to the AI. Some participants misinterpreted these arrows as indicating a future trend, rather than reflecting the AI's current weighting of the factors.

Overview

Table 1 *gives* a short summary of the work done in the phases.

Table 1. Summary of the different phases of the innovation turbine.

Phase 1	**Ideation** • Conducted literature research to scan the scientific horizon. • Held expert interviews with airline pilots to identify needs, generate ideas, and define user requirements for AI-based systems in the flight deck. • Developed initial ideas for an AI-based assistance system during an internal team workshop. • Formulated key concepts and ideas as the foundation for further development.
Phase 2.1	**System Analysis and System Model for the high-level system IPAS** • Conducted expert interviews with airline pilots, aeronautical engineers, and psychologists, along with literature research on decision-making processes, flight deck systems, explainable AI, and explainable interfaces. • Modelled a System of Systems. • Defined detailed use cases. • Designed the system model for the IPAS. • Conducted a user study with a basic prototype to identify explainability needs and additional user requirements. • Developed a concept for achieving operational explainability.
Phase 2.2	**System Analysis and System Model for the subsystem AICIS** • Defined detailed functionalities for the AICIS subsystem. • Specified concrete explainability features to implement operational explainability. • Developed a workflow diagram for AICIS using Rasmussen's ladder and Neisser's perceptual circle. • Created a detailed system model for the AICIS subsystem.
Phase 3	**Explorative Design, Development and Testing** • Developed a swim lane diagram to define individual AICIS interface pages and assign functionalities. • Iteratively designed the user interface, starting with sketches. • Implemented the demonstrator in a simulator environment. • Conducted a user study with the demonstrator to verify use cases, identify usability issues and focus particularly on the explainability of AI-based results. • Next step: Revise the demonstrator based on the study results and conduct further studies to validate the explainability concept and investigate the impact of using AICIS on decision making.

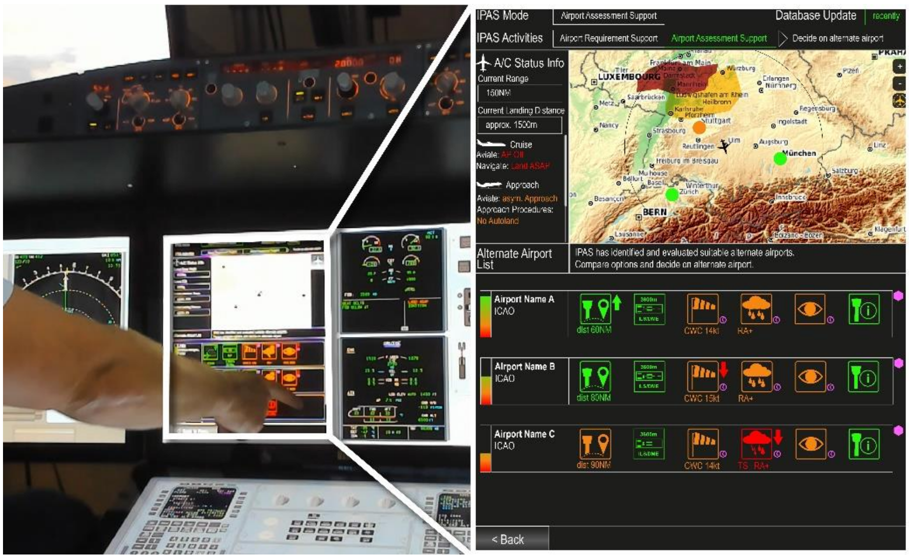

Figure 9. The working demonstrator, implementing into the simulator environment.

The results of this study will inform further revisions of the AICIS demonstrator, which will be iteratively developed and refined. This ongoing process will include additional studies with airline pilots, focusing on experimental and significant tests, such as the validation of the explainability concept and an investigation into the crew's decision-making performance with and without AICIS.

Discussion

At the beginning of the development, it was decided that the AICIS should be dedicated to researching and demonstrating AI explainability tailored to the needs, tasks and goals of airline pilots for future flight deck systems. It was unclear how to achieve this, as there was a scientific need for research and a regulatory requirement from the EASA roadmap for end-user-centred AI explainability, but no concrete implementation idea or approach. An exploratory approach was therefore considered appropriate. This methodology allowed for flexibility and adaptability during the development process, which was critical in addressing the unknowns surrounding the realisation of end-user explainability. The HSE paradigm and the associated Innovation Turbine served as effective methods to focus on exploration and innovation and proved beneficial in navigating the complexities of the project. By combining the Innovation Turbine with the Future Flight Deck Technologies framework, a structured approach was created for each phase of the Turbine to guide the development process while maintaining its exploratory nature. The synchronisation points, represented visually in the innovation turbine as Mach disks, were valuable because they helped to structure the output of the previous phase in terms of content and quality, and to form the transition to a new, expanding phase.

The explorative approach of the HSE paradigm proved useful and during the two and a half years of development and research, the benefits of this freedom were utilised and progress was made both in system development and in answering the research question. To approach the research goal, the combination of findings from the literature and the iterative and explorative development approach with trying out different ideas of what explanations could look like and what information could be needed, as well as conducting user studies with different demonstrators, proved to be effective. The studies explored how pilots interpret AI results and what information they need to understand AI-based information, leading to the identification of recurring patterns that indicate the information and explanation needed to achieve the required AI explainability. Based on this, a concept for operational AI explainability tailored to airline pilots and a specific use case was developed, which now needs to be tested in an experimental study. In addition to the research question, the explorative user studies identified many aspects of usability improvement, discovered potential for additional system functionality and use cases, and identified several aspects that now require structured experiments with a clear hypothesis for further investigation. One example of this is the different perception of the colour 'red' by the pilots. Some pilots identified information highlighted in red as 'out of limits', while others would interpret the colour as values that the AI considered 'bad', but not necessarily 'out of limits'. In addition, aspects were discovered by chance that were not predicted in advance. For example, in a study where two pilots used the IPAS as a crew, an improvement in collaborative decision making was observed. This observation is now to be investigated in a structured way.

The HSE paradigm can be useful in a scientific context where system development and research are carried out in parallel. For the development of the AICIS, the approach proved to be useful in the context of a doctoral thesis, where time constraints exist and structure was desired, but no research approach was available at the beginning and the potential of the given freedom should be fully exploited. Exploration leads to unexplored terrain, which may reveal possible research approaches. However, the HSE paradigm seems to be useful for any kind of innovation based on visions and ideas, even if it is not about exploring a research question. It should be noted that this discussion does not compare different approaches, but rather presents a single development approach and its application for a specific use case. Therefore, similar results can also be achieved with other approaches.

Conclusion

This paper presents how the HSE paradigm was adapted for the development of the IPAS and the subsystem AICIS in combination with a framework for user-centred design of future flight deck systems. The HSE paradigm is based on the metaphor of spatial exploration, and provides a descriptive model for what the explorative development of a new system looks like when exploration and innovation is desired in all phases of development. The methods used in the different phases of development and some of the results are presented, and finally it is discussed whether the approach was useful and what results could be achieved with it. The system was developed from scratch into a functioning demonstrator, and at the same time the research question of AI explainability tailored to airline pilots was addressed, with initial results already achieved using the explorative nature of the HSE paradigm. The findings from the explorative phase now need to be tested in structured experiments.

References

Banks, V.A., Plant, K.L. & Stanton, N.A. (2018). Development of a usability evaluation framework for the flight deck. In R. Charles and J. Wilkinson (Eds.), *Contemporary Ergonomics and Human Factor*. United Kingdome: CIEHF.

Banks, V.A., Plant, K.L., & Stanton, N.A. (2020). Leaps and shunts: designing pilot decision aids on the flight deck using Rasmussen's ladder. In R. Charles and D. Golightly (Eds.), *Contemporary Ergonomics and Human Factor*. United Kingdome: CIEHF.

Banks, V.A., Allison, C.K., Plant, K.L., Parnell, K.J., & Stanton, N.A. (2021). Using the Perceptual Cycle Model and Schema World Action Research Method to generate design requirements for new avionic systems. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *31*, 66-75.

Booher, H. R. (2003). *Handbook of human systems* integration (Vol. 23). USA: John Wiley & Sons.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies: why and how. In Proceedings of the 1st international conference on Intelligent user interfaces (pp. 193-200).

Djartov, B., Mostaghim, S., Papenfuß, A. & Wies, M. (2022). Description and First Evaluation of an Approach for a Pilot Decision Support System Based on Multi-attribute Decision Making. In *2022 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 141-147). IEEE.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

Endsley, M.R. (2023). Supporting Human-AI Teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior, 140*, 107574.

European Union Aviation Safety Agency (2023). *EASA-AI-Roadmap 2.0: A human-centric approach to AI in aviation.* Cologne, Germany: EASA.

Flemisch, F., Semling, C., Heesen, M., Meier, S., Baltzer, M., Krasni, A., & Schieben, A. (2013). Towards a balanced human systems integration beyond time and space: exploroscopes for a structured exploration of human–machine design spaces. In *HFM-231 SYMPOSIUM On "Beyond Time and Space"; NATO-STO Human Factors and Medicin Panel*.

Flemisch, F., Baltzer, M.C.A., Sadeghian, S., Meyer, R., Hernández, D.L. & Baier, R. (2019). Making HSI More Intelligent: Human Systems Exploration Versus Experiment for the Integration of Humans and Artificial Cognitive Systems. In W. Karwowski and T. Ahram (Eds.), *Intelligent Human Systems Integration 2019: Proceedings of the 2nd International Conference on Intelligent Human Systems Integration (IHSI 2019): Integrating People and Intelligent Systems,* (pp. 563-569). San Diego, California, USA: Springer International Publishing.

Flemisch, F., Preutenborbeck, M., Baltzer, M., Wasser, J., Kehl, C. & Grünwald, R. (2022). Human Systems Exploration for Ideation and Innovation in Potentially Disruptive Defense and Security Systems. In G. Adlakha-Hutcheon and A.

Masys (Eds.), *Disruption, Ideation and Innovation for Defence and Security* (pp. 79-117). Cham, Switzerland: Springer International Publishing.

Gunning, D., Vorm, E., Wang, J. Y. & Turek, M. (2021): DARPA's explainable AI (XAI) program: A retrospective. *Applied AI Letters*, 2. USA: John Wiley & Sons Ltd.

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. *Frontiers in Computer Science*, *5*, 1096257.

International Organization Standardization ISO (2019): ISO 9241-210:2019, Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems.

Kirwan, B., Charles, R., Jones, K., Li, W.C., Page, J., Tutton, W., & Bettignies Thiebaux, B. (2020). The human dimension in tomorrow's aviation systems. CIEHF-white paper.

Lee, J.D. & See, K.A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors, 46*, 50-80.

Lewis, J. R., & Sauro, J. (2021). Usability and user experience: Design and evaluation. Handbook of human factors and ergonomics, 972-1015.

Lucero, A. (2015). Using affinity diagrams to evaluate interactive prototypes. In Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part II 15 (pp. 231-248). Springer International Publishing.

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J.C., Parli, V., Shoham, Y., Wald, R., Clark, J. & Perrault, R. (2023). Artificial Intelligence Index Report.

Miller, M., Holley, S. & Halawi, L. (2023). The Evolution of AI on the Commercial Flight Deck: Finding Balance Between Efficiency and Safety While Maintaining the Integrity of Operator Trust. In *Artificial Intelligence, Social Computing and Wearable Technologies.* USA: AHFE International.

National Academies of Sciences, Engineering, & Medicine (2022). *Human-AI Teaming: State-of-the-Art and Research Needs.* Washington, DC, USA: The National Academies Press.

Parasuraman, R., Sheridan, T.B. & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*, 286-297. IEEE.

Parnell, K.J., Wynne, R., Plant, K.L., Stanton, N.A. (2021). User-centered design and evaluation of future flight deck technologies. In R. Charles and D. Golightly (Eds.), *Ergonomics and Human Factors*. United Kingdome: CIEHF.

Perri, L. (2023) What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle, accessed: 23th February 2024.

Plant, K. & Parnell, K. & Banks, V. & Wynne, R. & Stanton, N. (2023). Human Factors on the Flight Deck: A Practical Guide for Design, Modelling and Evaluation. 10.1201/9781003384465.

Rasmussen, J. (1974). The human data processor as a system component. Bits and pieces of a model. *Risø-M No. 1722*. Denmark: Risø National Laboratory.

Ribeiro, M.T., Singh, S. & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD* *international conference on knowledge discovery and data mining* (pp. 1135-1144). New York, NY, USA: Association for Computing Machinery.

Santos, I. S., Andrade, R. M., & Santos Neto, P. A. (2015). Templates for textual use cases of software product lines: results from a systematic mapping study and a controlled experiment. *Journal of Software Engineering Research and Development*, *3*, 1-29.

Save, L., Feuerberg, B. & Avia, E. (2012). Designing human-automation interaction: a new level of automation taxonomy. *Human Factors of Systems and Technology*.

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2017). Human factors methods: a practical guide for engineering and design. CRC Press.

Stefani, T., Deligiannaki, F., Berro, C., Jameel, M., Hunger, R., Bruder, C. & Krüger, T. (2023). Applying the Assessment List for Trustworthy Artificial Intelligence on the development of AI supported Air Traffic Controller Operations. In *2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC)* (pp. 1-9). IEEE.

Usai, M., Wasser, J., Herzberger, N. & Flemisch, F. (2023). Workshop on Participatory Exploration and Design of Human-Vehicle Interaction. In *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 275-278).

Würfel, J., Djartov, B., Papenfuß, A. & Wies, M. (2023). Intelligent Pilot Advisory System: The journey from ideation to an early system design of an AI-based decision support system for airline flight decks. In *Human Factors in Transportation.* USA: AHFE International.

Würfel, J., Papenfuß, A., Wies, M. (2024). Operationalizing AI Explainability Using Interpretability Cues in the Cockpit: Insights from User-Centered Development of the Intelligent Pilot Advisory System (IPAS). In: Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI. HCII 2024. Lecture Notes in Computer Science, vol 14734. Springer, Cham. https://doi.org/10.1007/978-3-031-60606-9_17