Insights from a Simulator Study: How the Intelligent Pilot Advisory System (IPAS) Could Impact Decision-Making in Flight Operations

Jakob Würfel[1,2](✉) and Ramón Lenting[3]

[1] German Aerospace Center (DLR e.V.), Lilienthalplatz 7, 38108 Braunschweig, Germany
[2] RWTH Aachen University, Aachen, Germany
jakob.wuerfel@dlr.de
[3] European Air Transport GmbH (EAT), August-Euler-Straße 1, 04435 Leipzig, Germany

Abstract. The Intelligent Pilot Advisory System (IPAS), developed by the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt (DLR)), is an AI-based cockpit assistance system that supports decision-making in normal, abnormal, and emergency flight situations. Previous work on IPAS has mainly addressed conceptual, technical, or explainability aspects and has not examined its operational impact. It is still unknown how automating parts of the pilots' decision-making process affects the decision-making structure, crew collaboration, information usage, and how the crews perceive the assistance in terms of usefulness and operational AI explainability. This exploratory study looked at six flight crews performing simulator flights involving technical malfunctions requiring diversions, both with and without the IPAS. The IPAS assists the crew by automatically identifying, assessing and selecting alternative airports based on technical, operational and environmental criteria. The crews' decision-making processes were evaluated through analysis of decision-making duration, expert reviews conducted by an experienced pilot, and questionnaires on system usefulness and explainability. The results indicated improvements in decision-making and collaboration within the crew when using the IPAS. Expert reviews revealed a slight tendency towards uncritical acceptance of the system's recommendations, suggesting a potential risk of automation bias. This study provides the first operational insights into the integration of the IPAS into the flight deck and its effect on decision-making. Further research with detailed and objective metrics is needed to comprehensively assess the effect of the IPAS on decision-making.

Keywords: AI-Based Cockpit Assistance · Operational AI Explainability · Flight Crew Decision-Making

W. -C. Li and A. Plioutsias (Eds.): HCII 2026, LNAI 16708, pp. 320–339, 2026.
https://doi.org/10.1007/978-3-032-29459-3_21

1 Introduction

The integration of Artificial Intelligence (AI)-based decision-support systems into the flight deck environment offers significant potential for supporting flight crews in complex and time-critical situations. Recent advances in machine learning and data-driven methods have enabled the development of AI systems that can process large amounts of operational data and provide recommendations to human operators. In aviation, such systems are increasingly discussed not only as automation tools, but as operational assistants that support human decision-making while keeping the human in the loop (e.g. [13,26,38]). Recognising AI's opportunities and risks in safety-critical domains, the European Union Aviation Safety Agency (EASA)'s AI-Roadmap 2.0 guides the integration of AI into aviation, highlighting human-centred integration, Human-AI teaming and operational explainability to secure trust, situation awareness, usability and safety [15].

The Intelligent Pilot Advisory System (IPAS) is being developed to investigate and demonstrate the introduction, operation, and human interaction with AI-based systems in the flight deck. The IPAS is an AI-based cockpit assistance system that supports flight crews during in-flight decision-making, particularly in abnormal and emergency situations. One functionality of the IPAS is its alternate airport support function, which generates, assesses, and presents potential alternate airport options based on operational constraints, aircraft state, environmental conditions, and company requirements. Previous work on the IPAS has focused on its conceptual design [32,34], technical architecture [4,5], and operational explainability [36].

Previous studies focusing on identifying explainability requirements involved only individual pilots. While the IPAS automates key elements of the decision-making process, its effect on flight-crew decision-making, perceived usefulness, and the operational explainability of the AI has not yet been evaluated with more realistic two-person cockpit crews. The aim of this exploratory study is to gain initial insights and identify potential risks, such as automation bias, before specific hypotheses are formulated and tested. Consequently, the study investigates how the IPAS influences two-person crew decision-making in abnormal scenarios, focusing on decision-making behaviour, information usage, and temporal structure, as well as perceived explainability and usefulness, to uncover possible advantages, effects, and risks associated with the system. As an exploratory investigation, no causal claims are made.

2 Motivation

On 15 November 2025, an Air Canada Rouge aircraft was on a flight from Antigua to Toronto. Forty-five minutes into the flight, the pilots detected a strong odour of fumes on board the Airbus A319 and decided to abort the flight for technical reasons, diverting to Punta Cana. One hour later the aircraft landed safely [25]. According to EUROCONTROL, approximately 30000 flights (0.32%

of all flights in Europe) landed at a different airport than originally planned in 2024 [12]. Diversions are required for a variety of reasons: weather, technical problems, medical emergencies, air-traffic constraints, or airport-related issues [29]. For each of these events a decision must be made, requiring good crew collaboration and communication, most likely using a structured decision-making model such as FORDEC (Facts, Options, Risks & Benefits, Decision, Execution, Check).

2.1 Decision-Making Processes in Airline Operations

The responsibility for the decision whether to divert and which alternate airport to select lies with the flight crew [16]. The application of Crew Resource Management (CRM) and structured decision-making processes are important elements in handling such scenarios safely and efficiently. CRM originated in the late 1970s as a response to crew-related accident causal factors, with the aim of reducing flight crew errors by improving non-technical skills such as communication, teamwork, leadership, and workload management [22]. Today, CRM training is widely implemented across airlines and is a standard safety component of aviation training programs (e.g., [14,19]). A central aspect within CRM is collaborative decision-making among flight crew members. One of the structured approaches to support this process is the FORDEC model [24,30]. Following to this model, crews systematically go through the FORDEC phases to gather facts, identify possible options, assess the associated risks and benefits, and come to an informed decision.

2.2 Challenges of Operational AI in Aviation

Since the 1980s the growing automation of commercial aircraft cockpits has turned the pilot from a manual operator into a system manager [2], exposing the "irony of automation" [1]. While automation raises efficiency and safety, it also creates human-factors risks such as out-of-the-loop loss of situation awareness, skill degradation, and automation bias [9,14,33]. Billings therefore advocated a human-centred automation approach that tailors automated functions to operators' capabilities [2].

The introduction of AI-based systems introduces additional challenges. Most AI algorithms behave as "black boxes" for humans, making their inner workings difficult or impossible to comprehend [18,21]. This opacity can result in poor situation awareness and increase out-of-the-loop effects [10]. Consequently, human-centred AI and Human-AI teaming demand explainable outputs that reveal the reasoning behind recommendations [8,11,21]. Another important aspect is calibrated trust. Users should neither distrust the system, which reduces acceptance, nor place too much trust in it, which can cause complacency, automation bias and unsafe overreliance [7,8,18,28].

The EASA published the AI Roadmap 2.0 as guidance for the introduction of AI in aviation [15]. In this roadmap, AI-based systems are classified according to their level of automation. The spectrum ranges from Level 1A systems, which provide automation support for information acquisition while full authority and responsibility remain with the end user, to Level 2B human-AI teaming systems. The latter involve the supervision of automatic decision-making and action implementation, with authority being shared between the human operator and the AI system. Beyond these levels, the roadmap also outlines the longer-term objective of Level 3A systems, characterised by safeguarded automatic decision and action implementation with limited authority remaining with the end user. To enable this, the roadmap defines so-called building blocks. One of these building blocks addresses human factors, identifying operational AI explainability as a central objective. Operational explainability refers to the explainability provided to the end user during system operation. This enables the user to understand the reasoning behind the system's output and interact with the AI system appropriately. EASA explicitly states that operational explainability is a key requirement for AI-based systems at all EASA AI levels [15,17].

3 Methodology

The aim of the study is to explore how the IPAS influences crew decision-making in a two-person cockpit and how pilots perceive its usefulness and explainability.

3.1 Description of the Intelligent Pilot Advisory System (IPAS)

The IPAS is an example of how AI-based decision support could be integrated into the flight deck. It covers use cases in normal operations [31,32] as well as use cases in abnormal and emergency situations [6,34]. Figure 1 provides an overview of the structure and functions of the IPAS. The IPAS consists of the AI Core Module (AICOM) and the AI-Crew Interaction System (AICIS) The present study focuses on the *Alternate Airport Support* function, which supports the Facts, Options, Risks & Benefits (FOR) phases of the FORDEC model, helping crews identify and evaluate suitable alternate airports in case of a diversion.

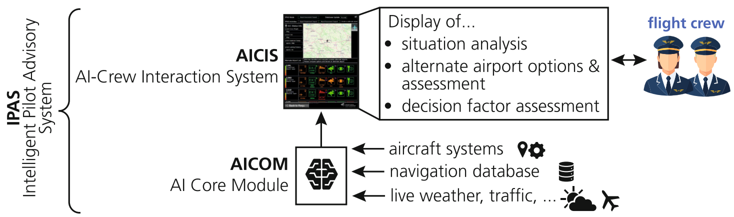

Fig. 1. Overview of the IPAS and its subsystems (Figure created by authors).

The AICOM collects information from the aircraft systems (position, technical status), accesses the navigation database (route, way-points, airports) and uses dynamic information such as current weather and traffic. The AICOM analyses the situation and, in the case of a required diversion, processes suitable alternate airport options by assessing nearby airports based on the detected situation and relevant decision factors. These airport options and their detailed assessments based on individual decision factors are presented to the pilots via the AICIS. The AICIS was developed with a focus on operational AI explainability, enabling crews to understand the rationale behind AI-based selections and assessments [36]. It was developed through a user-centred, iterative human-system exploration design process involving pilots from the start via interviews and mock-up studies [35], ensuring that the explanations and interface structure, a crucial element of the explainability concept, match the information and functionality required by the flight crew. Based on its current state of development, the IPAS can be classified as a Level 1B system according to the categorization of the EASA Roadmap 2.0 [15]: the IPAS is a cognitive assistance system that supports the crew in decision-making and selection tasks. In contrast, a Level 2 or 3 system would take over the selection and execution of the chosen alternate airport autonomously, leaving only supervisory oversight to the pilots. By remaining at Level 1B, the IPAS preserves full authority and responsibility with the crew.

A detailed description of the functionalities, the system structure and internal mechanisms of the IPAS is provided in [6].

The IPAS Used in the Conducted Study. For each FOR phase, the IPAS provides the information required and AI-based information for the respective step of the decision-making process. In the Facts phase, the system presents a situation analysis, the operational consequences of the situation, and the requirements for the search for a suitable airport. In the Options phase, suitable options are displayed with a preliminary assessment. For the Risks & Benefits phase, the crew can inspect assessed decision factors for each proposed option in order to support and justify the decision. The display used in the study is shown as follows: The *Alternate Airport Page* for the two scenarios used in the study, namely Spain and Italy, is shown in Fig. 2 and an excerpt from the *Detailed Airport View* of Rome is shown in Fig. 3.

On the *Alternate Airport Page* (representing Options), the current aircraft position is highlighted on a map and the proposed alternate airports are marked ①. The *Alternate Airport List* presents the available options (② shows the Spain scenario) in a ranked order according to the AI-based assessment result. A coloured bar visualizes the overall AI score for each option. In addition, the list provides the most relevant decision factors grouped into six categories (Position/Distance, Runway, Wind, Weather, Visibility, and Airport Operations) together with their respective assessments. The colours green, amber, and red indicate the influence of these factors on the airport assessment ③. ④ shows the *Alternate Airport List* for the Italy scenario. By selecting Rome, the *Detailed Airport View* (representing Risks & Benefits), shown in Fig. 3, opens, in which all

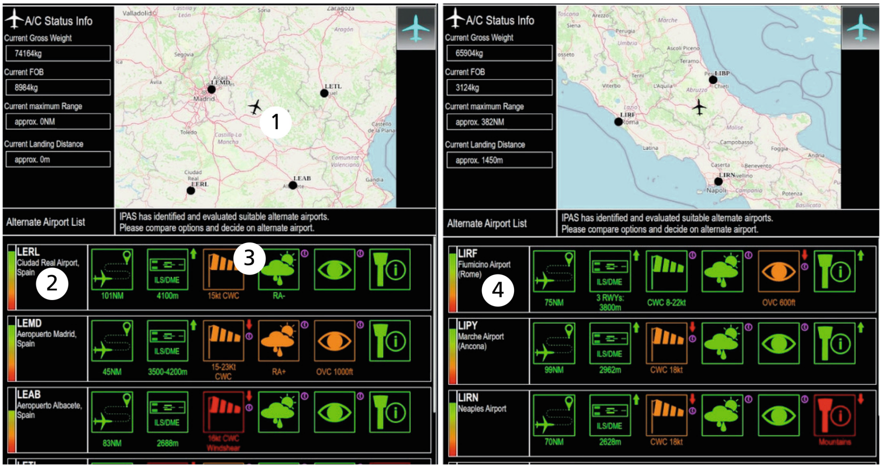

Fig. 2. *Alternate Airport Page* in the Spain and Italy scenario (Figure created by authors). ① Map, ② Alternate Airports Spain Scenario, ③ Factors, ④ Alternate Airports Italy Scenario.

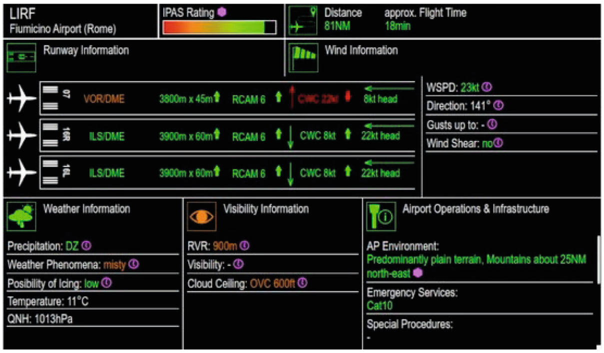

Fig. 3. *Detailed Airport View* for Rome Airport in the Italy scenario (Figure created by authors).

six categories are listed and a detailed overview of all decision factors influencing the airport assessment is provided.

For the purposes of this study, the information and assessments shown on the interface were generated with a Wizard-of-Oz paradigm [3], using simulated and predefined AI-Core outputs. As the two subsystems (AICOM and AICIS) had not yet been integrated and were still being developed separately, the mission effects, airport requirements, airport ratings and decision-factor assessments were all generated using the same multi-criteria decision-making approach underlying AICOM. None of these originated from a machine-learning model.

3.2 Study Design: Scenarios and Experimental Setup

Two simulator scenarios with technical malfunctions were used: an engine-fire in the Spain scenario and a fuel-leak in the Italy scenario. In both cases the crew had to initiate a diversion and select a suitable alternate airport under time pressure and challenging conditions (heavy rain, strong cross-winds, reduced visibility, limited maintenance, passenger-handling constraints, runway infrastructure, and approach procedures). Two different scenarios were selected to enable more flights to be carried out and evaluated with the available crews. The scenarios differ in terms of their environmental characteristics, such as geography and weather, but the chosen technical failures also require different responses, as fuel leaks are significantly more time-critical than (extinguished) engine fires. An experienced airline pilot simulated Air Traffic Control (ATC), providing realistic clearances and weather information. The flight deck environment was provided by the iSim flat-panel simulator of the Institute of Flight Guidance (DLR, Braunschweig), which was configured as an Airbus A321 cockpit (see [35] for details). The IPAS was mounted on the centre console.

Figure 4 shows the procedure of the study. Each crew flew both scenarios twice. These repetitions were chosen to establish comparability between the "with-IPAS" and "without-IPAS" settings. It was acknowledged that this may introduce training effects, but these were accepted given the exploratory nature of the study. At the beginning of the study, the participants received an initial briefing in which the purpose and procedure of the study were explained. As the participants were already familiar with the simulator from a previous study, no training session was necessary. Before each flight, the crew additionally received a short scenario briefing covering the planned route and basic flight parameters in order to facilitate immersion into the scenario. In the first two flights, the crews operated the scenarios without the IPAS under conventional settings (CONV) and had to solve the scenarios using their standard tools, ATC and procedures only. In the third and fourth flights, the crews operated the same scenarios with IPAS support. Finally, the participants answered questionnaires on the usefulness and explainability of the IPAS.

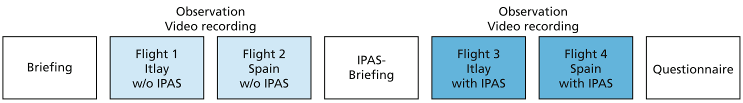

Fig. 4. Study procedure (Figure created by authors).

The study was approved by the DLR Ethics Committee on 30 January 2024. All participants signed an informed-consent form together with a GDPR-compliant data-protection declaration before the first briefing. The experiment reported here was one of several that were carried out consecutively on the same day. The whole test day lasted approx. 8h, during which participants moved from one study to the next. The present study required 1–1.5h. Each participant received a monetary compensation.

3.3 Pilot Sample

The study involved six flight crews, twelve commercial airline pilots in total. All participants held a valid Airbus type rating covering the A300, A320, A330, and A350 aircraft families. Age and number of flight hours were captured in categories: Age categories: $<$ 30 years; 30–39 years; 40–49 years; 50–59 years; $>$59 years; Total flight-hour categories: 0–1500 h; 1501–3000 h; 3001–5000 h; 5001–10000 h; $>$10,000 h. Three participants were younger than 30 years, seven were between 30 and 39 years old, and two were older than 40 years. Three participants had less than 3000 total flight hours, while nine participants had more than 3000 flight hours. The sample included four captains or senior first officers and eight first officers.

3.4 Evaluation Measures

The evaluation combined qualitative observations of crew decision-making, quantitative decision-making parameters, and questionnaires on explainability and system usefulness. An experienced airline pilot and flight instructor (more than 3500 flight hours on various aircraft types) observed the crews providing free text notes focusing on decision-making behaviour, information usage and crew interaction, and provided qualitative comments on decision-making quality. Two quantitative measures complemented the qualitative analysis: (i) the time required for the FOR phases and the total decision time, and (ii) the number of alternate airports mentioned and the subset that were considered in detail. To assess perceived explainability, the Explanation Satisfaction Scale (ESS) [23] was used. Additionally the IPAS were assessed by the participants using a tailor-made questionnaire addressing perceived usefulness, ease of use, stress, and perceived safety impact. The metrics and scales used can be found in the results Sect. 4.4.

4 Results

In total, 24 simulator flights were analysed. Twelve flights were conducted under conventional setting (CONV), and twelve flights were conducted with IPAS support, with six flights per scenario and setting. The scenarios are analysed and compared separately, as they are not comparable with each other.

4.1 Quantitative Decision Process Analysis

Table 1 presents the time required each FOR phase and the total decision time for the Italy and Spain scenarios, comparing the conventional and IPAS settings. The times shown are the means across the six crews in decimal minutes (standard

Table 1. Mean decision-making time (in minutes) for the FOR-phases in the Italy and Spain scenarios, averaged across all crews and compared between the CONV and IPAS settings.

Scenario	Setting		Facts	Options	Risks & Benefits	Total
Italy	CONV	min (SD) perc.	0.29 (0.22) 6.35%	3.23 (0.36) 70.30%	1.07 (0.64) 23.35%	4.59 (0.94)
	IPAS	min (SD) perc.	1.39 (0.67) 33.42%	1.22 (1.02) 29.47%	1.54 (0.97) 37.11%	4.15 (1.80)
		Δ (rounded)	+27%	-41%	+14%	-0.44
Spain	CONV	min (SD) perc.	0.90 (0.86) 10.01%	4.49 (0.99) 50.11%	3.58 (1.80) 39.88%	8.96 (2.68)
	IPAS	min (SD)	1.26 (0.72) 23.06%	1.27 (0.47) 23.21%	2.94 (1.48) 53.73%	5.47 (1.96)
		Δ (rounded)	+13%	-27%	+14%	-3.49

deviations are given in brackets). The Δ-row shows the rounded percentage change from CONV to IPAS for every phase and for the overall decision time.

For both scenarios the IPAS reduced the relative share of the Options phase (-41% in Italy, -27% in Spain) and increased the share of the Facts and Risks & Benefits phases. Mean total decision time dropped slightly from 4.59 min (CONV) to 4.15 min (IPAS) in Italy and from 8.96 min to 5.47 min in Spain. The standard-deviation values reveal that the decision times under the IPAS condition in the Italy scenario are considerably more variable (SD = 1.80 min) than under the conventional setting (SD = 0.94 min).

Table 2 shows the number of airports mentioned, as well as the number of airports in closer consideration per crew in the different settings in the Italy scenario. Table 4 shows the crews' final decision in the Italy scenario. Table 3 and Table 5 show the airport-related statistics for the Spain scenario.

4.2 Qualitative Observations on Decision-Making and Collaboration

Across both scenarios, qualitative expert observation revealed differences between the conventional and the IPAS-supported settings in decision-making structure, information usage, and crew interaction. To summarize the free text notes delivered by the flight instructor, they were post-hoc grouped into thematic clusters by the authors.

Structure: In the conventional setting, the decision-making process was frequently unstructured or incomplete. Several crews consulted the FOR phases in the wrong order, and some phases were skipped completely or addressed only briefly. When the IPAS was used, all crews followed a more structured process, consistently applying the FORDEC model. **Airports assessed:** In both settings some crews showed a tendency to lock onto the nearest or largest airport. In the IPAS setting, some crews considered only the first IPAS-recommended option. **Detail of assessment:** In the conventional setting, decision factors such as passenger handling, technical ground support, and company preference were often

Table 2. Airport statistics in the Italy scenario for each setting and crew.

Crew	Setting	Airports mentioned	Airports on closer consideration
1	CONV	4	2
	IPAS	4	1
2	CONV	3	2
	IPAS	3	1
3	CONV	3	3
	IPAS	4	3
4	CONV	5	1
	IPAS	4	2
5	CONV	2	2
	IPAS	4	2
6	CONV	6	2
	IPAS	4	2
Total Mean	CONV	3.83	2.00
	IPAS	3.83	1.83

Table 3. Airport statistic in the Spain scenario.

Crew	Setting	Airports mentioned	Airports on closer consideration
1	CONV	7	3
	IPAS	4	2
2	CONV	5	1
	IPAS	3	2
3	CONV	4	3
	IPAS	4	3
4	CONV	3	1
	IPAS	4	2
5	CONV	3	3
	IPAS	4	2
6	CONV	6	2
	IPAS	3	2
Total Mean	CONV	4.67	2.17
	IPAS	3.67	2.17

Table 4. Airports selection in the Italy scenario.

Airports	CONV	IPAS
Rome (LIRF)	4	5
Ancona (LIPY)	1	1
Naples (LIRN)	1	0
Pescara (LIBP)	0	0

Table 5. Airports selection in the Spain scenario.

Airports	CONV	IPAS
Ciudad (LERL)	0	1
Madrid (LEMD)	6	5
Albacete (LEAB)	0	0
Teruel (LETL)	0	0

omitted. With the IPAS, most crews evaluated a broader set of decision factors (weather, runway length, terrain, visibility, wind, passenger handling, technical support, and company preference) and assessed them in greater detail. **CRM:** Under the conventional setting, CRM issues were observed for one crew, including directive communication and unilateral decisions without a joint airport assessment in both scenarios. In the IPAS setting, the same crew showed a more structured and communicative interaction, exchanging preferences and conducting a joint assessment of options. **Overall**, the qualitative observations suggest that the IPAS was associated with more structured decision-making processes, more comprehensive use of available information, and more coordinated crew interaction. Notably, various decision-making patterns were observed: sequential option review, pairwise comparisons, comparing several alternatives against a favored candidate, generating additional options after an initial set, or producing the full set of alternatives initially.

Because the experiment employed a fixed order (CONV always before IPAS) and a within-subject design, the observed differences may be influences by order, practice, or training effects.

4.3 Explanation Satisfaction Scale

Table 6 shows the ESS results. The table shows the mean scores on the original 1–5 Likert scale (Strongly disagree (1) to Strongly agree (5)) as well as the Standard Deviation.

Table 6. Explanation Satisfaction Scale (ESS) results. (N=12)

Item	Mean (SD)
From the explanation, I know how the IPAS works.	3.83 (0.94)
The explanation of how the IPAS works is satisfying.	3.50 (0.80)
The explanation of how the IPAS works has sufficient detail.	3.25 (1.06)
The explanation of how the IPAS works seems complete.	3.08 (0.79)
The explanation of how the IPAS works tells me how to use it.	3.67 (0.98)
The explanation of how the IPAS works is useful to my goals.	4.08 (0.79)
The explanation of the IPAS shows me how accurate the IPAS is.	3.31 (1.00)
Total mean	3.53

The ESS results ratings with mean scores ranging from 3.08 to 4.08. The highest scores were observed for perceived usefulness of the explanations and for understanding how the system works, whereas lower scores were obtained for perceived completeness and level of detail.

4.4 Tailor Made Questionnaire

The results of the tailor-made questionnaire are presented in two tables. Table 7 summarizes the crew's overall attitude toward the IPAS concept on a seven-point Likert scale.

Table 7. Rating of tailor-made questionnaire. 7-point Likert scale (1 = Totaly dislike, 7 = Totaly like). Additionally total mean and standard deviation.

Item	1	2	3	4	5	6	7	Total mean (SD)
How much do you like the concept of the IPAS?	0	0	0	1	1	4	6	6.25 (0.97)

The concept of the IPAS was rated positively, with 11 out of 12 participants indicating that they liked the concept. The overall attitude toward the IPAS was very homogeneous: the total mean of 6.25 on the 7-point scale is accompanied by a low standard deviation (SD = 0.97), indicating that almost all participants rated the concept positively.

Table 8. Ratings of tailor-made questionnaire. 5-point Likert scale (1 = Strongly disagree, 5 = Strongly agree). Additionally Total mean and standard deviation.

Item	1	2	3	4	5	Total mean (SD)
Was it easy to use the IPAS considering it was only shortly introduced to you?	0	0	0	8	4	4.33 (0.49)
Does the use of the IPAS make your job more comfortable?	0	0	2	6	4	4.17 (0.72)
Does the IPAS provide you with a better overview of the options compared to when you collect the data conventionally?	0	0	3	6	3	4.00 (0.74)
Did you perceive a lower stress level when using the IPAS?	0	0	0	3	9	4,25 (0,45)
Do you think the use of an AI tool like the IPAS improves safety?	0	0	3	6	3	4.00 (0.74)

Table 8 reports more detailed statements about the usefulness of the IPAS on a five-point Likert scale. For each item the numbers in the table indicate the count of participants who selected the corresponding Likert point.

All participants reported that the system was easy to use despite only a brief introduction. Regarding perceived usefulness, 10 out of 12 participants indicated that a system like the IPAS could make their job more comfortable. 9 out of 12 participants indicated that the IPAS provides a better overview of the available options compared with conventional data collection, while 3 out of 12 responded neutrally. All participants reported a reduction in perceived stress when using the IPAS. With respect to safety, 9 out of 12 participants indicated that the use of an AI tool such as the IPAS could improve safety, while 3 out of 12 responded neutrally.

In total, the detailed statements yielded total means between 4.00 and 4.33 on the 5-point scale, with standard deviations ranging from 0.45 to 0.74, which reflects a consistently narrow spread of responses across the crew.

5 Discussion

This study provides exploratory insights into the impact of the IPAS on flight-crew decision-making, pilots' perceived usefulness of the system, and the explainability of its AI-based outputs. Potential implications for flight safety and operational efficiency are also discussed, followed by a brief outline of the study's limitations.

5.1 Effects on Flight Crew Decision-Making

The Use of the IPAS Seems to Primarily Affects the Temporal Structure of Decision-Making Rather than the Overall Time Needed. Due to the study design, this observation may be attributable to the study design, which means that no clear statement can be made. Also, a reduction in decision time does not imply higher decision quality, so no direct conclusions about decision quality can be drawn from timing alone. Across both scenarios, a consistent pattern emerges in which the relative time allocation shifts from the Options phase toward the Risks & Benefits phases when using the IPAS (-41% Options to +27% Facts and +14% Risks & Benefits in Italy scenario, -27% Options to +13% Facts and +14% Risks & Benefits in Spain scenario). This shift can be explained by the fact that the IPAS takes over the generation and rating of suitable options, as well as information gathering and assessment of relevant decision factors, thereby freeing cognitive and temporal resources for a more detailed assessment of the recommended option. The qualitative observations confirm this interpretation: crews examined each option in greater detail and evaluated a broader set of decision factors. A similar timely shift is observed towards the Facts phase. Qualitative observations confirm, that this phase was either skipped or only briefly addressed, whereas the dedicated Facts page of the IPAS encourages crews to explicitly review the situation, constraints, and requirements for potential airports, as can be confirmed by the observation.

The IPAS Does Not Change the Size of the Option Space, But it Could Influences How Options are Generated. With respect to the number of airports mentioned or taken into closer consideration, no effect of the IPAS was observed. In some cases, only a single option was examined in more detail when using the IPAS, which may indicate a risk of premature narrowing of the option space and neglected consideration of lower-ranked alternatives. This may indicate a tendency toward overtrust in the system. On the other hand, this could also be due to learning effects in the repetition of the scenario, whereby the pilots already had their preferred option in mind. In other cases, crews that considered only a single option under conventional settings compared multiple alternatives when using the IPAS, suggesting that the system can also support broader option consideration. Notably, in the Spanish scenario, most crews chose a different airport to the top-ranked IPAS recommendation, instead selecting the second-ranked option, which was also chosen by all crews under conventional settings. This suggests that crews do not follow the AI ranking blindly but actively assess and, where appropriate, overrule the recommendations based on their own operational judgement. At the same time, the experience from the first flight without IPAS may also have influenced the choice here. Taken together, risks such as overtrust, premature narrowing of the option space, and loss of critical thinking, known as automation bias, should be considered both in the design of AI-based systems and in related training and operational procedures. However, the study design does not permit the distinction between a narrowing of the option space caused by the system and a narrowing that occurs after participants have explored the scenario once.

It Seems that the IPAS is Contributing to a More Structured and Informed Decision-Making Process. The qualitative observations, supported by the qualitative analysis, indicate that the IPAS supports a more consistent application of the FOR phases. In several conventional runs, the Facts phase was incomplete or omitted, whereas with the IPAS the phases tend to be applied more consistently. The interface structure mirrors the FOR process, which appears to support both procedural guidance and flexible movement between phases, for example by allowing crews to return from detailed airport. Observations suggest that option assessments were often superficial and omitted relevant decision factors under conventional settings. In contrast, using the IPAS crews evaluated a broader set of decision factors, resulting in a more detailed airport assessment. This can be explained by the reduced effort required for collecting and analysing information, as well as by the overview of important factors and their colour-coded assessment provided by the IPAS. Even in scenarios where time savings are marginal, decisions made with IPAS support appear to be more informed. Standard practices primarily focused on checking the weather conditions at the next major airport (e.g., Madrid or Rome) and occasionally one additional option. However, other critical parameters, such as runway condition, technical support availability, passenger handling capabilities, and company preferences, were often neglected. This can be attributed to IPAS's detailed analysis of all relevant factors and alternatives, encompassing a wider range of decision factors than typically considered in conventional decision-making processes.

The IPAS May Influence Decision Strategies and Recurring Decision Patterns. Across crews, the observation revealed recurring decision-making patterns, such as sequential option generation after an initial assessment, early fixation on a single option, or systematic one-by-one comparison of alternatives. The IPAS may influence not only decision-making structure and information use but also decision strategy. This points to a promising direction for future research, for example along the lines of decision pattern analyses such as those reported by [37].

Observation Suggests that the IPAS May have a Positive Impact on CRM. Finally, a remarkable but anecdotal observation was that the CRM issues previously observed in one crew under conventional settings did not occur when the IPAS was used. The IPAS may serve as an third neutral agent in the cockpit hierarchy, providing an objective baseline that mitigates interpersonal friction and appears to facilitate coordinated discussion, flatten hierarchies, and support joint decision-making. This interpretation remains speculative and would require a detailed investigation. However, it is very difficult to reproduce CRM problems specifically for a study.

5.2 Explainability and Perceived Usefulness

The Explainability of the IPAS is Rated Positively and as Helpful for Decision-Making. The explainability of the IPAS received an overall ESS mean of 3.55, indicating that participants generally agreed that the explanations were useful and helped them understand how the system works, while still leaving room for improvement. In particular, the explanations provided by the IPAS, such as the presentation of relevant decision factors influencing airport ratings, the use of color coding to indicate individual factor assessments and the indication of factor weightings [36], supported pilots in understanding the justification behind alternate airport assessments. These explanations were also perceived as useful for decision-making, indicating that they supported the operational task.

There Could be a Tension Field Between Depth of Explanation and Operational Applicability. At the same time, the perceived level of detail and completeness of the explanations received lower ratings. Taken together, this suggests a trade-off between the amount and depth of explanatory information on the one hand and the need for explanations at an appropriate level of detail that allows pilots to focus on operational decision-making on the other, as also discussed in the literature (e.g., [7,21]). In this tension field, the IPAS appears to provide fewer but operationally relevant explanations, which were sufficient to support decision-making without overloading the crew, while at the same time possibly providing slightly too little detail. This tension field should be carefully balanced when designing explainable and operational AI-based systems.

The IPAS is Perceived as a Useful and Supportive Decision-Support Tool. The results of the tailor-made questionnaire further indicate that the IPAS was perceived as useful and supportive for the decision-making process. The concept was positively received, the system was considered easy to use, and its use was associated with increased perceived comfort and reduced perceived stress. In addition, a majority of pilots indicated that the IPAS provides a better overview of available options compared to conventional data collection.

5.3 Potential Implications for Flight Safety and Operational Efficiency

When evaluating socio-technical systems, consideration should be given not only to the system qualities discussed in this study, but also to their overall impact on organisation, society and the environment [20]. While the present study was not designed to assess safety outcomes or operational performance metrics directly, the observed changes in flight crew decision-making processes suggest potential implications for flight safety, operational efficiency and sustainability. From a flight safety perspective, the IPAS was associated with a more structured and explicit application of the FORDEC model. Research in human factors and CRM has repeatedly shown that incomplete situation assessment, premature fixation on a single option, and unstructured trade-offs are recurrent contributors to inadequate operational decisions [22,27,30]. These patterns were frequently observed under conventional settings, whereas the IPAS-supported setting was characterized by more systematic consideration of operational constraints, explicit articulation of consequences, and more coordinated crew interaction. Although no causal claims can be made, these mechanisms are widely regarded as supporting factors for safe flight operations and align with established principles of structured aeronautical decision-making [24]. In addition to safety-related considerations, the findings indicate possible implications for operational efficiency. Diversion decisions affect flight-path length, holding time, airport suitability and secondary-diversion risk. The observed shift in time allocation from option generation toward information assessment and comparison suggests that the IPAS may reduce the effort required for ad hoc information gathering and processing while enabling more informed assessments of available alternatives. From an airline operations and management perspective, more informed and timely diversion decisions may contribute to reduced downstream effects such as extended fuel burn, additional ground handling complexity, and passenger disruption. Reduced fuel burn and emissions are a likely indirect benefit. These potential effects should be quantified by integrating decision-making analyses with operational, economic and environmental performance models.

5.4 Limitations and Future Research Directions

This study has several limitations that need to be considered when interpreting the results. First, the study followed a non-randomized within-subject design with a fixed order of settings and scenarios. This limits the ability to attribute observed effects unambiguously to the use of the IPAS and does not allow for strong causal claims. Future studies should employ randomized and counterbalanced designs to control for learning and order effects. Second, the qualitative evaluation of decision-making was conducted by a single expert observer. While this provided rich domain-specific insights, it also introduces subjectivity and potential bias. If qualitative observations are used as a metric in future work, multiple domain experts should be consulted to increase the objectivity of qualitative observations. The present study included quantitative performance measures (e.g., the duration of the FOR phases). The qualitative observations were used to complement these metrics by assessing how crews used information, interacted, and by revealing decision-making patterns. Information that is not evident from timing data alone. A promising direction for future research is to quantify the qualitative findings, for example by analysing phase-transitions and information-usage, to obtain a deeper and more objective insight into the decision-making structure. As previously mentioned, the impact of the IPAS could also be evaluated through an analysis of decision-making patterns and strategies that differ between conventional and IPAS-supported settings (e.g., [37]).

With respect to further development of the IPAS, future work should focus on the tension between explainability and information load, for example by enabling adaptive levels of explanation detail depending on the operational context and user needs. In addition, training concepts and procedural guidance should be developed to support critical reflection when working with AI-based systems and to mitigate risks of automation bias. Future research could investigate how implementing IPAS translates into measurable outcomes relating to safety, efficiency, costs and emissions, for example by combining simulator studies with operational data, fuel burn models and airline performance metrics.

6 Conclusion

This paper presented an exploratory simulator study on the effects of the IPAS on flight-crew decision-making and on pilots' perceptions of its explainability and usefulness. The results suggest that the IPAS primarily supports a more structured and more detailed consideration of facts, possible options, and their risks and benefits. The study revealed a shift from time spent on searching for options towards time spent on the detailed assessment of the decision factors and available options when using the IPAS, potentially leading to more informed decisions. Pilots perceived the IPAS as useful and supportive, and they rated its explanations as helpful for understanding and justifying system outputs. At the same time, the findings point to a potential tension between the depth of explanations and the cognitive and temporal resources required to process them, indicating that the level of explanatory detail can directly influence operational applicability. Importantly, CRM-related issues observed in the conventional condition disappeared when the IPAS was used, suggesting that the system can act as a neutral third-party that facilitates coordinated crew interaction. At the same time, the risk of premature narrowing of the option space and over-trust (related to automation bias) was noted.

Overall, the results highlight the potential of user-centred, explainable AI systems such as the IPAS to support flight-crew decision-making in complex operational contexts, while underlining the importance of careful system design, appropriate explainability, and training to mitigate known risks of operational AI: premature narrowing of the option space, automation bias, and overreliance on AI-generated rankings. The findings suggest that AI-based decision-support systems such as the IPAS can influence decision-making structure, information usage and CRM, with possible impacts on flight safety, operational efficiency and sustainability.

Acknowledgments. 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. Bainbridge, L.: Ironies of automation. Automatica **19**(6), 775–779 (1983). https://doi.org/10.1016/0005-1098(83)90046-8
2. Billings, C.E.: Human-centered aviation automation: Principles and guidelines. Technical Memorandum NASA-TM-110381. NASA Ames Research Center, Moffett Field, CA, USA (1996)
3. Dahlbäck, N., Jönsson, A., Ahrenberg, L.: Wizard of oz studies - why and how. Knowl.-Based Syst. **6**(4), 258–266 (1993). https://doi.org/10.1016/0950-7051(93)90017-N

4. Djartov, B., Mostaghim, S., Papenfuß, A., Wies, M.: Description and first evaluation of an approach for a pilot decision support system based on multi-attribute decision making. In: IEEE Symposium Series on Computational Intelligence, pp. 141–147 (2022). https://doi.org/10.1109/SSCI51031.2022.10022076
5. Djartov, B., Mostaghim, S., Papenfuß, A., Wies, M.: A learning classifier system approach to time-critical decision-making in dynamic alternate airport selection. In: IEEE Congress on Evolutionary Computation (CEC), pp. 1–8 (2024). https://doi.org/10.1109/CEC60901.2024.10612016
6. Djartov, B., Würfel, J.: Navigating Decisions in the Cockpit: The Intelligent Pilot Advisory System, pp. 305–325. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-83512-4_18
7. Droogenbroeck, C., Rankova, E., Papenfuss, A., Bos, T., Zon, R.: Human-ai teaming - challenges from a practitioner's perspective. In: Harris, D., Li, WC. (eds.) Engineering Psychology and Cognitive Ergonomics, pp. 337–354. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-93718-7_22
8. Dubey, A., Abhinav, K., Jain, S., Arora, V., Puttaveerana, A.: Haco: A framework for developing human-ai teaming. In: Proceedings of the 13th Innovations in Software Engineering Conference (Formerly Known as India Software Engineering Conference). ISEC '20 (2020). https://doi.org/10.1145/3385032.3385044
9. Endsley, M.R.: Automation and situation awareness. In: Proceedings of the 1996 International Conference on Human-Computer Interaction (1996). https://api.semanticscholar.org/CorpusID:261054213
10. Endsley, M.R.: Autonomous horizons: System autonomy in the air force – a path to the future. volume i: Human-autonomyteaming (2015). https://doi.org/10.13140/RG.2.1.1164.2003
11. Endsley, M.R.: Supporting human-ai teams:transparency, explainability, and situation awareness. Comput. Hum. Behav. **140**, 107574 (2023). https://doi.org/10.1016/j.chb.2022.107574
12. EUROCONTROL: Cost of diversion (2025). https://ansperformance.eu/economics/cba/standard-inputs/latest/chapters/cost_of_diversion.html
13. Commission, E.: Just a rather very intelligent system (Jarvis) (2026). https://doi.org/10.3030/101114692
14. European Union Aviation Safety Agency: Crm training implementation (2017). https://www.easa.europa.eu/en/document-library/general-publications/crm-training-implementation
15. European Union Aviation Safety Agency: Easa-ai-roadmap 2.0: A human-centric approach to ai in aviation (2023). https://www.easa.europa.eu/en/domains/research-innovation/ai
16. European Union Aviation Safety Agency: Easy access rules for air operations – nco.gen.105 pilot-in-command responsibilities and authority (2023). https://www.easa.europa.eu/en/document-library/easy-access-rules/online-publications/easy-access-rules-air-operations?kw=Journey&page=4
17. European Union Aviation Safety Agency: Easa concept paper: Guidance for levelâĂŕ1 &âĂŕ2 machine-learning applications (2024). https://www.easa.europa.eu/en/domains/research-innovation/ai
18. Ezer, N., Bruni, S., Cai, Y., Hepenstal, S.J., Miller, C.A., Schmorrow, D.D.: Trust engineering for human-ai teams. Proc. Human Factors Ergonomics Soc. Ann. Meet. **63**(1), 322–326 (2019). https://doi.org/10.1177/1071181319631264
19. Federal Aviation Administration: Advisory circular 120-51d: Crew resource management training (2001). https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-51D.pdf

20. Flemisch, F., Preutenborbeck, M., Herzberger, N., et al.: Vom Teufelsquadrat zum nachhaltigen Engelsdiamanten und holistischen Bow-Tie-Modell: Methodenentwicklung für die ganzheitliche Analyse, Gestaltung und Entwicklung von nachhaltigen Arbeitssystemen. Zeitschrift für Arbeitswissenschaft **78**, 146–159 (2024). https://doi.org/10.1007/s41449-024-00418-5, https://doi.org/10.1007/s41449-024-00418-5
21. Gunning, D., Vorm, E., Wang, J.Y., Turek, M.: Darpa's explainable ai (xai) program: A retrospective. Appl. AI Lett. **2**(4), (2021). https://doi.org/10.1002/ail2.61
22. Helmreich, R.L., Foushee, H.C.: ChapterâĂŕ1 – why crm? empirical and theoretical bases of human-factors training. In: Kanki, B.G., Helmreich, R.L., Anca, J. (eds.) Crew Resource Management (Second Edition), pp. 3–57. Academic Press, San Diego, second edn. (2010). https://doi.org/10.1016/B978-0-12-374946-8.10001-9
23. Hoffmann, R.R., Mueller, S.T., Klein, G., Litman, J.: Measures for explainable ai: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-ai performance. Front. Comput. Sci. **5** (2023). https://doi.org/10.3389/fcomp.2023.1096257
24. Hörmann, H.J.: For-dec - a prescriptive model for aeronautical decision making. In: 21st WEAAP Conference, Dublin (1994). https://elib.dlr.de/27044/
25. Hradecky, S.: Accident: Canada rouge a319 near punta cana on nov 15th 2025, fumes on board (2025). https://avherald.com/h?article=5307c73e
26. Miller, M., Holley, S., Halawi, L.: The evolution of ai on the commercial flight deck: Finding balance between efficiency and safety while maintaining the integrity of operator trust. In: Proceedings of the AHFE 2023 (2023). https://doi.org/10.54941/ahfe1004175
27. Orasanu, J., Martin, L., Davison, J.: Errors in aviation decision making: Bad decisions or bad luck? In: Fourth Conference on Naturalistic Decision Making. NASA-Ames Research Center, Warrenton, VA (1998). https://ntrs.nasa.gov/citations/20020063485
28. Parasuraman, R., Manzey, D.: Complacency and bias in human use of automation: An attentional integration. Human factors **52**, 381–410 (2010). https://doi.org/10.1177/0018720810376055
29. SKYbrary: Diversion (2025). https://skybrary.aero/articles/diversion
30. Soll, H., Proske, S., Hofinger, G., Steinhardt, G.: Decision-making tools for aeronautical teams: For-dec and beyond (2016). https://doi.org/10.1027/2192-0923/a000099
31. Ternus, S.: User-Centered Design of an Intelligent Pilot Advisory System for Non-Emergency Scenarios: A Case Study. Master's thesis, Hochschule Trier (2024). https://elib.dlr.de/206128/
32. Ternus, S., Würfel, J., Papenfuß, A., Wies, M., Rumpler, M.: Exploring functionalities for an intelligent pilot advisory system in normal operation. In: Engineering Psychology and Cognitive Ergonomics: 21st International Conference, EPCE 2024 (HCII 2024), pp. 235–247. Washington, DC, USA. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-60728-8_19
33. Wiener, E.L., Curry, R.E.: Flight-deck automation: promises and problems. Ergonomics **23**(10), 995–1011 (1980). https://doi.org/10.1080/00140138008924809
34. Würfel, J., Djartov, B., Papenfuß, A., Wies, M.: 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: AHFE 2023 (2023). https://doi.org/10.54941/ahfe1003844

35. Würfel, J., Flemisch, F.O.: Human system exploration for the ai-based flight deck decision support system ipas. In: de Waard, D., et al. (eds.) Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2024 Annual Conference. Avail (2024) https://doi.org/10.60575/31n5-2a06
36. Würfel, J., Papenfuß, A., Wies, M.: 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, pp. 297–315. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-60606-9_17
37. Zhang, Z.T., et al.: Beyond recommendations: From backward to forward ai support of pilots' decision-making process. Proc. ACM Human-Comput. Inter. **8**(CSCW2), 1–32 (2024). https://doi.org/10.1145/3687024
38. Zhang, Z.T., Liu, Y., Husmann, H.: Pilot attitudes toward ai in the cockpit: Implications for design. In: Nürnberger, A. (ed.) Proceedings of the 2021 IEEE International Conference on Human-Machine Systems, pp. 1–6. IEEE (2021). https://doi.org/10.1109/ICHMS53169.2021.9582448