Building Trust by Supporting Situation Awareness: Exploring Pilots' Design Requirements for Decision Support Tools

CARA STORATH*, fortiss GmbH, Research Institute of the Free State of Bavaria, Germany
ZELUN TONY ZHANG*, fortiss GmbH, Research Institute of the Free State of Bavaria, Germany
YUANTING LIU, fortiss GmbH, Research Institute of the Free State of Bavaria, Germany
HEINRICH HUSSMANN, LMU Munich, Germany

Supporting pilots with a decision support tool (DST) during high-workload scenarios is a promising and potentially very helpful application for AI in aviation. Nevertheless, design requirements and opportunities for trustworthy DSTs within the aviation domain have not been explored much in the scientific literature. To address this gap, we explore the decision-making process of pilots with respect to user requirements for the use case of diversions. We do so via two prototypes, each representing a role the AI could have in a DST: A) Unobtrusively hinting at data points the pilot should be aware of. B) Actively suggesting and ranking diversion options based on criteria the pilot has previously defined. Our work-in-progress feedback study reveals four preliminary main findings: 1) Pilots demand guaranteed trustworthiness of such a system and refuse trust calibration in the moment of emergency. 2) We may need to look beyond trust calibration for isolated decision points and rather design for the process leading to the decision. 3) An unobtrusive, augmenting AI seems to be preferred over an AI proposing and ranking diversion options at decision time. 4) Shifting the design goal toward supporting situation awareness rather than the decision itself may be a promising approach to increase trust and reliance.

CCS Concepts: • **Human-centered computing → Interaction design process and methods**.

Additional Key Words and Phrases: human-AI interaction, decision support tools, decision support systems, human-AI teaming, aviation

ACM Reference Format:
Cara Storath, Zelun Tony Zhang, Yuanting Liu, and Heinrich Hußmann. 2022. Building Trust by Supporting Situation Awareness: Exploring Pilots' Design Requirements for Decision Support Tools. In *CHI TRAIT '22: Workshop on Trust and Reliance in Human-AI Teams at CHI 2022, April 30, 2022, New Orleans, LA.* ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

AI-based *decision support tools* (DSTs) are an extensively researched topic within the HCI community [1, 3, 7, 14, 17, 21, 26]. Fast information processing in particular is a benefit that AI can contribute to the decision-making of human-AI teams [9]. In modern aviation, some of pilots' most demanding tasks are the assessment of abnormal and novel situations and decision-making in complex, uncertain, and equivocal environments [27]. Supporting pilots with a DST is therefore a promising application for AI in aviation, particularly in use cases where a large amount of information analysis is required in high-workload and time-critical contexts. One specific example for such a challenging situation is a

*Both authors contributed equally to this research.

diversion: In this case, an emergency or other abnormal event requires the crew to divert to another airport than the original destination.

Research on DSTs suggests that in order to build a beneficial and trustworthy DST, it is crucial to consider the context of use and embed the AI into the surrounding workflow [25]. Within the context of clinical decision-making, Yang et al. do so with an "unremarkable AI", meaning that the AI's functionality is to augment the decision-making process by displaying prognostics [24]. Van Berkel et al. discuss a similar concept under the term of "continuous" human-AI interaction, which they characterize as "interaction as commentary" [20]. To learn more about the context of diversion decisions as well as possibilities and requirements for AI-based DSTs in aviation, we follow an iterative *Research through Design* (RtD) approach, with the presented work as the first feedback loop. We built two DST prototypes, each representing a different role AI could have in such a system: A) Unobtrusively hinting at data points the pilot should be aware of. B) Actively suggesting and ranking diversion options based on criteria the pilot has previously defined. Our findings suggest that diversion decisions are rather processes than isolated decision points. Consequently, the main value of DSTs might rather be supporting pilots' situation awareness than the moment of the decision itself. Additionally, pilots demand guaranteed trustworthiness of the AI and reject trust calibration via explainability during the emergency. These requirements may be better fulfilled by an AI which continuously produces hints in an unobtrusive way, rather than by one that is actively suggesting options at decision time.

This workshop paper makes two contributions. First, we present two new interaction concepts for DSTs, especially adding to design concepts for unremarkable AI. Second, we give an industry perspective based on a work-in-progress study, showing potential design requirements and challenges regarding trust and reliance for DSTs in aviation, as well as deeper insights into the decision-making process of aircrews.

2 BACKGROUND

2.1 Decision-making in aviation

Aviation is a highly automated domain where most of the motoric and tactical tasks of flying are already partially or fully supported by automation [2]. For an uneventful flight, the automation can take over almost the entire flight from right after takeoff up until landing. Tasks left to the crew include operating and supervising the automation, communicating with air traffic control (ATC), and decision-making. These tasks are particularly important in an emergency or abnormal situation, which likely falls outside the competence of the automation. In such a situation, the human capability to assess novel situations and to solve problems is essential [27].

Decision-making in emergency or abnormal situations is one of the most challenging tasks for pilots, as a multitude of technical, operational, and environmental factors needs to be considered and prioritized. To make good decisions in such complex and dynamic situations, it is crucial for pilots to form and maintain *situation awareness* (SA). According to the commonly adopted model of Endsley [5], SA consists of three levels: 1) *perception* of elements in the current situation, 2) *comprehension* of the current situation, and 3) *projection* of the future status.

In our current work, we consider the issue of diversions, that is situations where an emergency or abnormal situation requires the crew to divert to an airport different from the original destination. In such a case, pilots need to first decide whether a diversion is necessary, and if so, which alternative destination to divert to. Possible reasons for a diversion include bad weather at the original destination, a technical failure, or a medical emergency among the passengers.

2.2 Decision support tools

The effects of AI-based DSTs on human decision-making are an increasingly popular subject for empirical research [13]. One commonly studied topic here is the issue of *trust calibration*: Decision-makers should have appropriate trust toward the AI support, which means they should neither rely on AI when it is wrong, nor ignore it when it is correct. Most pertaining studies investigate the effect on trust calibration created by explanations of AI suggestions [18, 22, 23, 26]. However, Lai et al. [13] point out a potential gap between the research findings and how DSTs are actually used in practice, as most studies focus on the moment of the decision and ignore factors like workflow or context of use. Yet, these factors can have a major impact on the effectiveness of DSTs. For example, research on clinical DSTs suggests that their main value may not be to support the moment of the decision itself, but rather other subtasks surrounding the decision [10, 11, 25]. Similarly, Lubars and Tan [15] emphasize the importance of defining a suitable task for the AI. In their survey, they find a strong preference among participants for machine-in-the-loop designs in which humans are leading the process. This corresponds to the interaction-as-commentary paradigm described by van Berkel et al. [20], in which the AI continuously processes input and reacts if necessary. This paradigm was represented in the form of an AI highlighting potential polyps with visual markers in a clinical inspection task [19].

Findings like these point toward the importance of a holistic view when designing DSTs—ignoring the context of decision-making might negatively impact trust and reliance. In healthcare for instance, research has already shown evidence that DSTs appear to be barely adopted despite their effectiveness in laboratory settings [4, 25].

3 METHOD

To inform our future work, we created two low-fidelity prototypes of an AI-based diversion assistance system and discussed them with professional pilots. In the following, we present our prototypes and the design rationale behind them, followed by a description of the study design.

3.1 Scenario

The scenario presented to the pilots was an in-flight diversion due to a medical emergency, in this case, a passenger having a heart attack. In this type of emergency, the pilot should land the aircraft soon to enable medical care. This most likely means diverting to another airport in case the planned destination is not close enough.

The prototypes illustrate a general diversion assistance system, which in theory should not only work for a diversion due to a medical emergency but also in case of other abnormal situations, such as technical failure. For the study however, only the specific user journey of a medical emergency was prototyped completely.

3.2 Prototypes

We built two click dummy prototypes for diversion assistance systems. They mainly differ in the role the AI plays within the system and how the pilot can interact with it.

3.2.1 Prototype A—Local Hints. In this prototype, the AI was designed to continuously evaluate and highlight decision-relevant information without suggesting particular diversion options (Fig. 1). This is done by displaying all surrounding airports and their information regarding certain criteria in a table. These criteria are important for diversions, such as flight time, runway length, weather at the destination, or distance to the next hospital. The airports are ordered from left to right according to a criterion specified by the pilot (e.g. flight time). Each table cell displays the value of one criterion for one airport. The AI's main task is to evaluate these values and to display warnings and alerts in case a

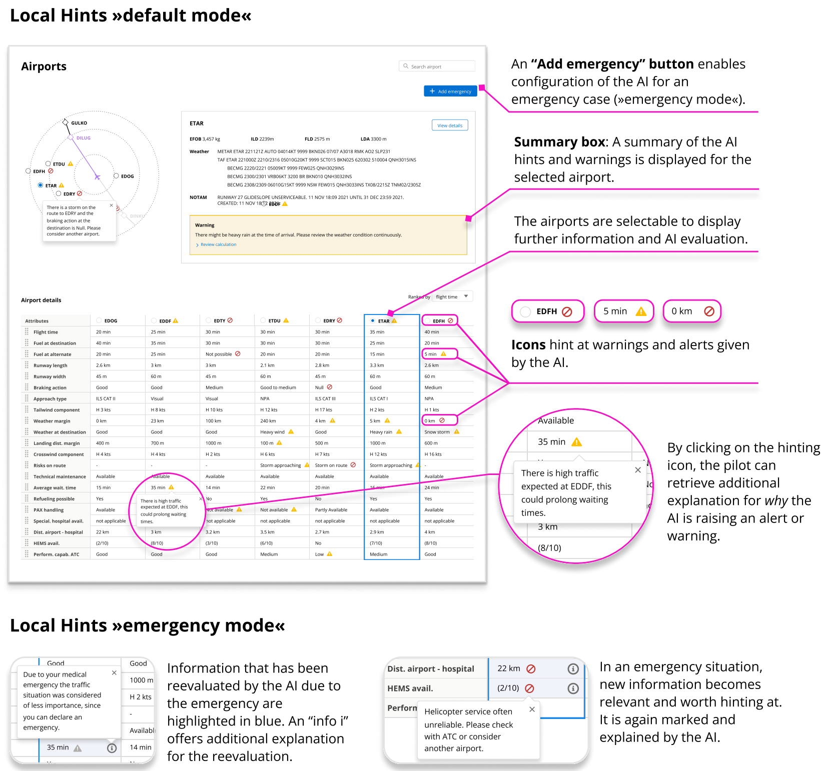

Fig. 1. **Prototype A—Local Hints:** The application can be used in »default mode« and »emergency mode«. Switching the mode changes the evaluation of the AI. Icons hint at criteria that the AI considers worth alerting or warning about. Further information is displayed on demand via popovers. The airports are displayed in the header of the table, each row displays the information regarding a certain criterion (listed on the left).

value requires extra attention. These warnings and alerts are displayed in the form of icons and are called "Local Hints". By clicking on the icon, the pilot can retrieve additional information as to why the AI considered the value hint-worthy. The pilot additionally has the option to select an airport to see a summary box. This summary box contains the main hints of the selected airport and serves as a short communication of the main points highlighted by the AI.

The prototype can be used in two modes: »default mode« and »emergency mode«. Within the »default mode«, the information is displayed and evaluated generally, with no particular reason for diversion in mind. This mode aims to provide SA, by giving the pilot an overview of the current situation and potential risks. In case an emergency occurs, the pilot can activate the »emergency mode« and enter a concrete reason for a diversion, in this case "medical emergency". The AI then reevaluates all the data and gives hints according to the new situation. For a medical emergency, this means for example displaying a warning if the helicopter service is predicted to be unavailable. In case the evaluation has changed for a certain criterion due to the emergency mode, the corresponding table cell is indicated visually.

3.2.2 Prototype B—Global Suggestions. In prototype B, the AI makes concrete suggestions of diversion options. In contrast to prototype A, the airports are ranked according to the AI's calculation of the most suitable diversion options (Fig. 2). Before getting a suggestion, the pilot needs to enter criteria for the AI suggestions. The criteria are the same as

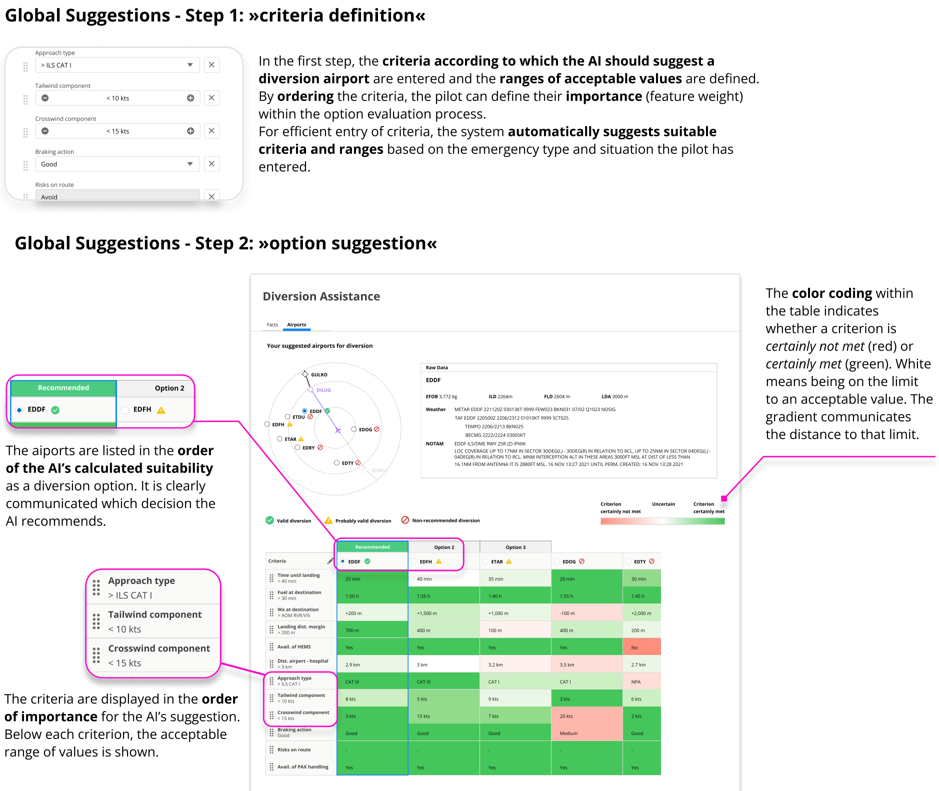

Fig. 2. **Prototype B—Global Suggestions:** In the first step, the pilot can define the criteria according to which the AI should calculate its suggestions for a diversion airport. In the second step, diversion options are displayed in the order of the calculated suitability. The color coding communicates how certain the AI is that a criterion will be met.

in prototype A. However, while prototype A always displays all criteria available to the AI, prototype B only considers the subset of the criteria defined by the pilot. The pilot can furthermore define the importance of the individual criteria and define an acceptable range of values for them. Suitable criteria and values are suggested by the AI based on the emergency situation entered in order to speed up the input process. Letting the pilot define and check the criteria first before displaying a suggestion aims to provide more control over the AI's evaluation. On the screen suggesting suitable diversion options, the criteria previously defined are again listed below the airports. A color coding communicates how likely the corresponding criterion will be met for the respective airport. The calculated values for the individual criteria together with their likeliness of fulfillment are meant to serve as an explanation for the AI's suggestions.

3.2.3 Design process. We created the described prototypes by following a loose RtD approach in which the first two authors explored different ways in which an AI could support a pilot in making diversion decisions. The interfaces matured by reflecting on the design decisions taken and giving feedback to each other. The design process was informed by several informal interviews with pilots and experts from the aviation industry as well as by an existing concept from prior work [6].

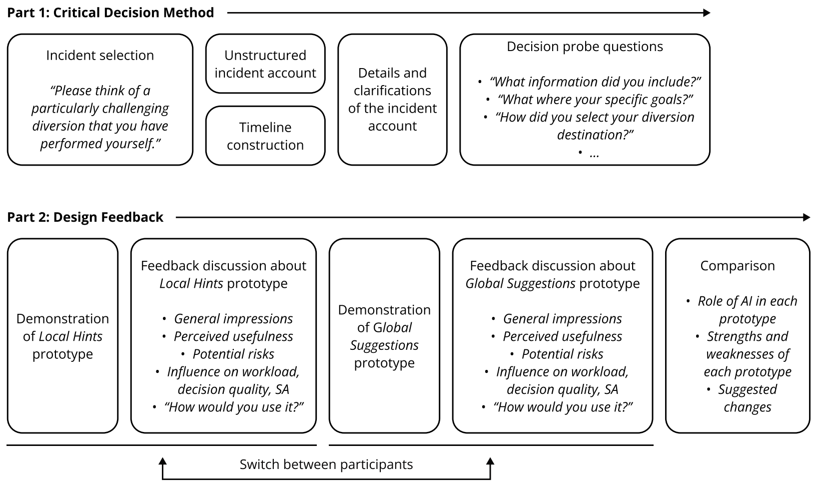

Fig. 3. Overview of the study sessions. Each session consisted of two parts: A short, half-hour interview about diversions according to the Critical Decision Method (Section 3.3.1), and a feedback discussion about our two prototypes (Section 3.3.2).

The prototypes were built with the design tool Figma[1] which also enables implementing click dummies. To speed up the design process, we partly used components of the Lightning Design System[2].

3.3 User Study

We have so far recruited and interviewed three experienced pilots (all male, average age: 41.3 years, average flying hours: 6,833 hours), with two of them being German airline pilots. The third one has a background as fighter pilot, but also regularly test-flies passenger aircraft. The pilots received no incentives for their participation. Generally, this is a target group which is very difficult to address. As of this writing, we are planning to recruit at least three more airline pilots for our study, so we are reporting on work in progress. We conducted the study over the video conferencing platform Webex[3] and recorded all sessions. Each session had a length of about two hours and consisted of two parts, as described in the following. Fig. 3 shows an overview of the sessions.

3.3.1 Critical Decision Method interview. We started each session with a half-hour semi-structured interview according to a pared-down version of the *Critical Decision Method* (CDM) [12], a method that is widely used to elicit domain expert knowledge about complex decision-making tasks. We asked participants to think about one particularly challenging diversion from their own experience. If a pilot had no personal diversion experience, we asked him to think of a relevant simulator training situation instead. When participants had chosen a case, we then asked for a brief description of the incident, from the moment the pilot became aware of the problem, until the completion of the diversion. While participants described their diversion, we simultaneously took note of their account in the form of a rough visual timeline using the collaboration tool MURAL[4]. After they had finished their account, we showed this timeline to

[1] https://www.figma.com
[2] https://www.lightningdesignsystem.com
[3] https://www.webex.com
[4] https://www.mural.co

participants via screen sharing so that they could clarify details or add missing pieces of information. Lastly, based on this account, we probed their decision-making for more details using probing questions taken from [12].

This CDM-based interview served two purposes: For one, we wanted to gain a deeper understanding of the operational complexities of a diversion decision. Additionally, the specific incidence from their own experience also served as a concrete example that participants could refer back to and elaborate on during the second part of the session.

3.3.2 Feedback discussion. In the second part of the sessions, we discussed our two prototypes with the participants, again in the form of semi-structured interviews. By confronting pilots with two different interfaces and ways of AI support, we aimed to get feedback on the pros and cons of each, learn more about potential design opportunities and challenges, and about decision-making in aviation in general. We did so by first discussing both prototypes separately before asking about them in comparison. Between participants, we switched the order of the prototypes to mitigate order effects. For each prototype, we first demonstrated it to participants through screen sharing by showcasing a typical user journey for our scenario while explaining the actions. Afterward, we asked participants for their feedback. Besides questions about first impressions or useful features and potential risks, we also asked about the influence of such a system on workload, decision quality, and SA in situations of differing degrees of risk and urgency. One of our particular concerns was that DSTs might produce inadequate outputs, for instance due to events that the system is not designed to account for. As a further probe, we therefore first asked participants to describe how they would use the system for a diversion decision. Following that, we asked how their usage would look like in a situation where an important factor is not considered by the system. For this, we let them assume that the destination appearing most favorable based on the AI outputs would be unavailable due to a no-fly zone caused by political disturbances.

After discussing both prototype variants in the described manner, we closed the interviews by asking participants how they perceived the role of the AI in both variants, letting them compare the strengths and weaknesses of each, and asking for suggested changes to the prototypes.

4 FINDINGS

Even though we have had only three participants so far, the discussions with the pilots have already produced interesting preliminary findings that we want to share with the community. In the current state of our work, these findings are based on a loose analysis of the interviews without the application of rigorous analysis methods.

4.1 Decision-making as a process instead of isolated decision points

Our participants' statements suggest that for pilots to be able to trust and rely on the diversion assistance, the system should fit into the whole process of decision-making, rather than treating diversions as isolated decision points. For instance, the airline pilots highlighted that diversions are very rare, so a system meant to be used only at the moment of the actual diversion decision would also be used very rarely. This would make it hard for pilots to familiarize themselves with the system and to rely on it in a critical situation: *"I think I have roughly experienced four or five diversions in twelve years. And if I only started using such a system once such a situation arises, I don't know if pilots would be open to it."* (P2)

Hence, a diversion assistance system needs to be integrated into pilots' overall workflow, as P2 elaborated: *"I need to use the system in everyday situations for me to also use it in abnormal situations."* A possible way to do so is to cater to pilots' need for maintaining SA. Pilots do not start the decision process only when a diversion becomes necessary, but *"always look into the future"* (P2) and try to evaluate potential landing options:

> *"Having a quick look, what's up with the airports around me, could I land there, you always do this. If I now had such a table, where I could see on the display, 'oh look, in Frankfurt, there is something with the NOTAMs*[5]*, they have closed it now,' or 'the braking coefficient went down there, we can't land there anymore.' This situational awareness is the key to safe flight operation."* (P3)

Always having a valid plan B ready is considered good airmanship and even required by some airlines. In the best case, pilots are therefore well prepared to make a diversion decision when necessary:

> *"In the Nice incident for example, there we didn't use anything at the moment of the decision, we just executed our plan. But with such a system, you could have dealt with the situation long, long in advance, that would have been great, that you could already say 'we can fly there and there, there it looks fine.' "* (P2)

P2 even imagined that a system similar to the *Local Hints* variant could be used to simulate possible scenarios as part of pre-planning: *"You could maybe just click through it for yourself, like 'I'll just do medical emergency now, what would happen then?' You could do this to get the big picture."*

4.2 Guaranteed trustworthiness instead of trust calibration

"So if I can trust the system ..." was a conditional clause used frequently by all three pilots. They all emphasized that being able to rely upon the system is a fundamental precondition for using and accepting the system in the first place. Reacting to the question whether he would not be worried that the system was not aware of some information or miscalculated anything, P1 stated: *"If I would use the system with this approach, I wouldn't use it at all. [...] Then it occupies me more than taking the decision myself."*

The pilots mentioned the need for trustworthiness in the context of completeness, recency, and validity of the information: P1 required the system to recalculate corresponding factors when conditions change, for instance if a mechanical error causes an increased fuel consumption. When being asked about his probable behavior in case he would know that the best-rated airport is within a no-fly zone, but the AI does not have access to that information, P2 answered: *"If you are designing a system which has so extensive influence on our decision-making, then it must be a system on which we can really rely that such things are recognized and processed."*

Even though the pilots highlighted the importance of reliability, the airline pilots P2 and P3 stressed that *"we never trust something blindly, we always check it for its validity or reliability"* (P2). P1 and P3 mentioned the potential issue of overtrust. P1 especially referred to the green color coding of the criteria in the *Global Suggestions* interface as being a potential source of overtrust, whereas P2 evaluated the green color coding positively: He mentioned that it would give him a *"good feeling"* seeing a lot of green for his chosen option. P3 worried about the complacency effect [16] and suggested that systems should be designed in a way that pilots *"still need to think along."*

4.3 Reliability of information instead of how AI uses the information

When it comes to assessing the reliability of the diversion assistance, our participants were most concerned about the reliability of the information displayed in the table, less about how exactly the algorithm uses the information for its outputs. This became apparent when comparing participants' reactions to the explanations in both prototypes. The visualization of the *Global Suggestions* prototype shows how the prioritization and likelihood of fulfillment of the criteria result in the diversion suggestions. The summary box of the *Local Hints* variant on the other hand does not tell anything about how the algorithm would work, but explains potential problems of the respective diversion option.

[5] *Notices to Air Missions* (originally Notices to Airmen): important real-time notices about abnormal status for persons involved in flight operations.

Still, P1 and P2 perceived the latter as more transparent (*"I think I definitely liked the other one better, that I have a summary up there"* (P1)). Only P3 saw an advantage with the *Global Suggestions* prototype for supervising the system. He considered it less likely to overlook inadequate AI behavior in that variant, though not because of the explanation visualization, but because *"you first select the things yourself and define them, actively."*

While no participant was interested in the inner workings of the AI, all of them emphasized that *"it must definitely be understandable where it gets its information from, how it gets its information"* (P2). P2 also suggested providing links from the processed data used by the AI to the raw data sources so that pilots could check the validity of the information. Preferably, these should be links to resources that pilots are already familiar with, like the briefing package:

> *"If you could say 'add to briefing package,' [...] that you could get an input from the AI, 'this is my thought because it says so here and there,' a link so to say from the AI suggestion, where does it take it from, where does it say so, how is the connection?"* (P2)

The priority of information reliability over the inner workings of the AI aligns with where our participants see the biggest value of a DST. In today's cockpits, information acquisition is the most laborious part of diversion decisions. Accordingly, the two airline pilots see the value more in the quick information access than the decision suggestion: *"I am confident that I could make similarly good decisions. [...] But the workload is significantly lower"* (P3).

4.4 Other findings

4.4.1 Legal aspects. Another insight which may have an impact on trust and reliance is that a certified DST could be used to defend a pilot's decision in front of a court. P3 explained:

> *"If something went to shit and you are standing in front of the judge, you can say, this is the screenshot and this is officially certified and that's what I've followed. Nowadays, these are things that you can't neglect. A lot of thoughts in flying are about 'is this legally okay?' Am I able to defend that in front of the prosecutor?"* (P3)

4.4.2 Additional criteria for diversions. Pilots mentioned four additional criteria that can play a role during a diversion: current passenger wellbeing (P2), familiarity with an airport (P2), medical equipment available at an airport (P2), and cost and effort for their employer (P1, P2, P3). P2 mentioned the current wellbeing of his passengers as a reason for not trying a second approach at a destination with bad weather: *"First, to land somewhere quickly, because the people weren't feeling well at all anymore with that weather."* The same pilot also elaborated that his familiarity with the chosen airport was a decisive factor in one case because he knew it had a supporting infrastructure. Personally knowing the airport was also brought up as a diversion criterion by two other pilots during our user research preceding the current study.

5 DISCUSSION

Taken together, our preliminary findings point to the importance of a wider view on the design space of AI-based DSTs in aviation. Specifically, a range of trust-related challenges and solutions appears to lie outside the moment of the decision itself and more within the process surrounding a decision. We elaborate on this insight in more detail in the following. Given the work-in-progress nature of our work and the low number of participants so far, our statements are not empirically validated, but they might provide the community with interesting food for thought nevertheless.

5.1 Opportunities beyond trust calibration for isolated decision points

Empirical research on AI-based DSTs predominantly focuses on supporting human decision-makers with AI-generated decision suggestions [13]. However, our participants' statements indicate that such a narrow focus on isolated decision points may be inadequate to successfully shape trust and reliance for diversion assistance systems. Similar to clinicians [8], pilots appear to reject the idea of calibrating their trust for every single decision, as they feel this would create more effort than making the decision without assistance. Moreover, pilots would likely be unwilling to rely on a system designed to be used exclusively during diversions, since such a system would be used too rarely to be familiar with it.

As an alternative to calibrating trust for isolated decision points, our findings suggest that a more unobtrusive, "commenting" AI [20] that integrates well into existing workflows might be a promising option, similar to what Yang et al. explored in a healthcare context [24]. This would allow pilots to build trust and to get a feeling of the capabilities and limits of the AI over time. For a diversion assistance system, this could mean focusing more on supporting information acquisition and SA rather than suggesting decisions.

This insight shows that finding a suitable role for the AI based on user feedback is crucial. Even though the objective for DSTs is to improve human decision-making, the most obvious approach of directly tackling this goal through decision suggestions might not always be the best option. Especially when the trustworthiness of AI-suggested decisions cannot be guaranteed, the role of the AI might need to be shifted to a more indirect form of decision support.

5.2 Design challenges for diversion assistance systems

Our participants' statements further hint at several design challenges related to trust and reliance, all of which are notably reaching beyond the moment of the decision and into the process around it.

5.2.1 Unobtrusive integration of AI. For pilots to get used to and hence to rely on a diversion assistance tool, the system needs to be well integrated into their familiar workflow and information sources. While our prototypes, especially the *Local Hints* variant, encompass ideas in that regard that are worth exploring, how exactly such an unobtrusive AI for diversion assistance could look like needs to be further investigated. A possible guiding principle for such a design is to aim for an AI that *"sits in the background, that [...] provides information and thinks along"* (P2).

5.2.2 Expectation management. As discussed in Section 5.1, finding an appropriate role for AI is crucial to build DSTs that pilots would be willing to rely on. Concomitant with the choice of the role of the AI is the proper communication of this role, and especially its limits, to manage expectations. Explanations could be useful in this regard, but the way the AI functionality is integrated into the overall workflow is likely just as important.

5.2.3 Influence of legal aspects. As touched on in Section 4.4.1, pilots might be incentivized to follow the AI when in doubt due to legal consequences. How this influences reliance in practice, in particular overreliance, and how it should be factored into the DST design might constitute interesting questions for further research.

5.2.4 Soft factors of decision-making. The complex decisions in aviation are further complicated by the influence of soft factors. Pilots' personal familiarity with an airport or passengers' wellbeing can tip the scales in favor of one or another diversion option. These are factors that are difficult for the system to take into account. Therefore, a trustworthy system needs to be flexible enough not to burden the pilot or pose a risk in case such unknown factors come into play. Designing such flexible human-AI interactions remains a challenge.

6 CONCLUSION

We have presented an in-progress study to understand design requirements and opportunities for trustworthy decision support tools for diversions. Our preliminary findings suggest that focusing on the point of the actual decision may be insufficient to shape trust in such AI-based systems. Instead, promising mechanisms for trust-building may be situated in the process surrounding a decision. Through the discussion of two prototypes, one with more active AI suggestions and one with a more unobtrusive AI, we find signs of benefits of the latter approach for trust and reliance.

We plan on expanding our study with additional participants to strengthen and extend our findings. Additionally, we intend to further explore interaction concepts for unobtrusive AI in aviation, especially with a focus on supporting situation awareness.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) under the LuFo VI-1 program, project KIEZ4-0, and by the Bavarian Ministry of Economic Affairs, Regional Development and Energy.

We thank our participants for their time and valuable input.

REFERENCES

[1] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the whole exceed its parts? The effect of AI explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, Yokohama, Japan, 81:1–81:16. https://doi.org/10.1145/3411764.3445717 .

[2] Charles E. Billings. 1996. *Human-centered aviation automation: principles and guidelines.* Technical Report NASA-TM-110381, A-961056, NAS 1.15:110381. NASA Ames Research Center, Moffett Field, CA United States. 222 pages. https://ntrs.nasa.gov/citations/19960016374 .

[3] Adrian Bussone, Simone Stumpf, and Dympna O'Sullivan. 2015. The role of explanations on trust and reliance in clinical decision support systems. In *Proceedings of the 2015 International Conference on Healthcare Informatics (ICHI 2015)*. IEEE, Dallas, TX, USA, 160–169. https://doi.org/10.1109/ICHI.2015.26 .

[4] Glyn Elwyn, Isabelle Scholl, Caroline Tietbohl, Mala Mann, Adrian G. K. Edwards, Catharine Clay, France Légaré, Trudy van der Weijden, Carmen L. Lewis, Richard M. Wexler, and Dominick L. Frosch. 2013. "Many miles to go …": a systematic review of the implementation of patient decision support interventions into routine clinical practice. *BMC Medical Informatics and Decision Making* 13, Suppl 2 (Nov. 2013), S14:1–S14:10. https://doi.org/10.1186/1472-6947-13-S2-S14 .

[5] Mica R. Endsley. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors* 37, 1 (March 1995), 32–64. https://doi.org/10.1518/001872095779049543 .

[6] Claudia Fernández Henning. 2021. *FOR-DEC and beyond – conceptual HMI design for a diversion assistance system.* Bachelor Thesis. Technische Hochschule Ingolstadt, Ingolstadt.

[7] Ben Green and Yiling Chen. 2019. The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 50:1–50:24. https://doi.org/10.1145/3359152 .

[8] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. Designing AI for trust and collaboration in time-constrained medical decisions: a sociotechnical lens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, Yokohama, Japan, 659:1–659:14. https://doi.org/10.1145/3411764.3445385 .

[9] Mohammad Hossein Jarrahi. 2018. Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. *Business Horizons* 61, 4 (July 2018), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007 .

[10] Annika Kaltenhauser, Verena Rheinstädter, Andreas Butz, and Dieter P. Wallach. 2020. "You have to piece the puzzle together": implications for designing decision support in intensive care. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*. ACM, Eindhoven, Netherlands, 1509–1522. https://doi.org/10.1145/3357236.3395436 .

[11] Kensaku Kawamoto, Caitlin A. Houlihan, E. Andrew Balas, and David F. Lobach. 2005. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ : British Medical Journal* 330, 7494 (March 2005), 1–8. https://doi.org/10.1136/bmj.38398.500764.8F .

[12] Gary A. Klein, Roberta Calderwood, and Donald MacGregor. 1989. Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics* 19, 3 (May 1989), 462–472. https://doi.org/10.1109/21.31053 .

[13] Vivian Lai, Chacha Chen, Q. Vera Liao, Alison Smith-Renner, and Chenhao Tan. 2021. Towards a science of human-ai decision making: a survey of empirical studies. *arXiv:2112.11471 [cs]* (Dec. 2021), 1–36. http://arxiv.org/abs/2112.11471 .

[14] Vivian Lai and Chenhao Tan. 2019. On human predictions with explanations and predictions of machine learning models: a case study on deception detection. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. ACM, Atlanta, GA, USA, 29–38. https://doi.org/10.1145/3287560.3287590 .

[15] Brian Lubars and Chenhao Tan. 2019. Ask not what AI can do, but what AI should do: towards a framework of task delegability. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS 2019)*. Curran Associates, Inc., Vancouver, Canada, 57–67. .

[16] Raja Parasuraman, Robert Molloy, and Indramani L. Singh. 1993. Performance consequences of automation-induced 'complacency'. *The International Journal of Aviation Psychology* 3, 1 (1993), 1–23. https://doi.org/10.1207/s15327108ijap0301_1 .

[17] Forough Poursabzi-Sangdeh, Daniel G. Goldstein, Jake M. Hofman, Jennifer Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, Yokohama, Japan, 237:1–237:52. https://doi.org/10.1145/3411764.3445315 .

[18] Philipp Schmidt and Felix Biessmann. 2020. Calibrating human-AI collaboration: impact of risk, ambiguity and transparency on algorithmic bias. In *Machine Learning and Knowledge Extraction (CD-MAKE 2020)*. Springer International Publishing, Dublin, Ireland, 431–449. https://doi.org/10.1007/978-3-030-57321-8_24 .

[19] Niels van Berkel, Omer Ahmad, Danail Stoyanov, Laurence Lovat, and Ann Blandford. 2020. Designing visual markers for continuous artificial intelligence support: A colonoscopy case study. *ACM Transactions on Computing for Healthcare* 2, 1 (2020), 7:1–7:24. https://doi.org/10.1145/3422156 .

[20] Niels van Berkel, Mikael B. Skov, and Jesper Kjeldskov. 2021. Human-AI interaction: intermittent, continuous, and proactive. *Interactions* 28, 6 (Nov. 2021), 67–71. https://doi.org/10.1145/3486941 .

[21] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland, UK, 601:1–601:15. https://doi.org/10.1145/3290605.3300831 .

[22] Xinru Wang and Ming Yin. 2021. Are explanations helpful? A comparative study of the effects of explanations in AI-assisted decision-making. In *Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21)*. ACM, College Station, TX, USA, 11. https://doi.org/10.1145/3397481.3450650 .

[23] Fumeng Yang, Zhuanyi Huang, Jean Scholtz, and Dustin L. Arendt. 2020. How do visual explanations foster end users' appropriate trust in machine learning?. In *Proceedings of the 25th International Conference on Intelligent User Interfaces (IUI '20)*. ACM, Cagliari, Italy, 189–201. https://doi.org/10.1145/3377325.3377480 .

[24] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: fitting intelligent decision support into critical, clinical decision-making processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, Glasgow, Scotland, UK, 1–11. https://doi.org/10.1145/3290605.3300468 .

[25] Qian Yang, John Zimmerman, Aaron Steinfeld, Lisa Carey, and James F. Antaki. 2016. Investigating the heart pump implant decision process: opportunities for decision support tools to help. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, San Jose, California, USA, 4477–4488. https://doi.org/10.1145/2858036.2858373 .

[26] Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. ACM, Barcelona, Spain, 295–305. https://doi.org/10.1145/3351095.3372852 .

[27] Zelun Tony Zhang, Yuanting Liu, and Heinrich Hußmann. 2021. Pilot attitudes toward AI in the cockpit: implications for design. In *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*. IEEE, Magdeburg, Germany, 1–6. https://doi.org/10.1109/ICHMS53169.2021.9582448 .