

Ergonomics

ISSN: 0014-0139 (Print) 1366-5847 (Online) Journal homepage: www.tandfonline.com/journals/terg20

Seeing inside the black box: Effects of algorithmic knowledge on system trustworthiness in aviation

Anneke Hamann, Alexander Heise, Helge Lenz & Anja Gampe

To cite this article: Anneke Hamann, Alexander Heise, Helge Lenz & Anja Gampe (20 Oct 2026): Seeing inside the black box: Effects of algorithmic knowledge on system trustworthiness in aviation, Ergonomics, DOI: 10.1080/00140139.2026.2738038

To link to this article: https://doi.org/10.1080/00140139.2026.2738038

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 20 Oct 2026.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=terg20

ERGONOMICS
https://doi.org/10.1080/00140139.2026.2738038

Taylor & Francis
by informa

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Seeing inside the black box: Effects of algorithmic knowledge on system trustworthiness in aviation

Anneke Hamann, Alexander Heise, Helge Lenz and Anja Gampe

German Aerospace Center, Institute of Flight Guidance, Braunschweig, Germany

ABSTRACT

Automation has shaped aircraft design for the past seven decades, moving from fully manual control to highly automated systems. Despite this, pilots retain final authority for safety-critical decisions, which requires continuous trust calibration. This study investigated how algorithmic transparency influences pilots' learned trust in an AI-based cockpit assistant. Twenty-one commercial pilots were randomly assigned to an early transparency condition (detailed explanations of the AI's algorithms and performance provided before the tutorial) or a late transparency condition (the same information delivered after the experiment). Each pilot used the assistant during a tutorial and a realistic simulated flight. Mixed-effects ordinal models showed higher trust ratings for the early compared to the late transparency group during both tutorial and flight, and a larger trust increase between phases. In conclusion, providing pilots with timely, clear information about AI processes modestly improves trust and supports calibrated reliance.

Practitioner Summary: Investigated whether giving pilots early algorithmic explanations of an AI cockpit assistant increases trust. Used a mixed-design experiment with 21 Airbus pilots, comparing early vs. late transparency across tutorial and flight phases. Found early transparency raised both initial and dynamic trust, especially ease-of-use.

ARTICLE HISTORY
Received 8 June 2026
Accepted 18 September 2026

KEYWORDS
Human automation teaming; transparency; algorithms; trust; aviation; pilots; artificial intelligence

Introduction

Automation in modern aircraft

The aviation domain, and especially aircraft design, has historically been shaped by automation. Less than 70 years passed between the early designs of fully manually controlled aircraft and automatic landing systems that needed no human intervention (Billings 1991). The current 'fourth generation glass cockpit' aircraft provide pilots with a very high level of automation. The on-board systems do not only alleviate the physical demands of manually controlling the aircraft but also provide assistance for cognitive tasks such as computations on fuel consumption, optimal flight level and weight distribution (Sheridan 1999). In short, automation can handle most of the aviation and navigation tasks, which mainly leaves the pilots communicating, making decisions and monitoring the automation (Sheridan 1999). Yet, pilots remain a critical factor in ensuring the safety of their aircraft, crew and passengers (Taylor and Cotter 2020). They have the final authority over cockpit decisions and bear the responsibility for the outcomes of the actions taken based on the decisions (Holford 2022). Thus, pilots undergo extensive and recurrent simulator training to be able to identify and solve technical problems with their aircraft under time pressure and adapt their flight strategy based on unexpected circumstances (Feary 2018; Wickens and Dehais 2019).

The emergence of artificial intelligence

The rapid diffusion of artificial intelligence (AI) is reshaping virtually every facet of modern life, from personal assistants and recommendation systems to complex decision-support tools in safety-critical domains (Candrian and Scherer 2022). While AI is not yet present in aircraft, the European Union Aviation Safety Agency (EASA) is currently working on frameworks for AI assistance and certification regulations (EASA 2023; Gajduskova et al. 2025). This transformation promises gains, but the success of AI-enabled technologies hinges on a human-centred prerequisite (Billings 1991): users must trust the systems that they are asked to rely on. Trust is understood not merely as an affective attitude

CONTACT Anja Gampe anja.gampe@dlr.de German Aerospace Center, Institute of Flight Guidance, Braunschweig, Germany
© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

but as a multidimensional construct that is associated with workload, situation awareness, and ultimately performance (Hoff and Bashir 2015; Lee and See 2004; Mayer, Davis, and Schoorman 1995; Muir 1994).

Today's aircraft systems are deterministic, i.e. they follow rules that can be formulated and documented in handbooks and can be learned and looked up. AI systems, however, do not necessarily follow explicit rules and may provide analyses and recommendations based on computations that are not obvious to human eyes (Carabantes 2020). In short, AI systems are mostly black boxes. If a pilot does not fully understand how an AI system has come to a certain result and cannot verify it based on the available information, it stands to reason that their trust in the system plays an important role in accepting or disregarding its assistance. In fact, a pilot's willingness to heed AI recommendations can be critical (Halawi, Miller, and Holley 2024; Hernandez, Ayo, and Panagiotakopoulos 2021). Empirical investigations have shown that pilots' trust in AI-driven decision aids is strongly modulated by the clarity of the system's reasoning (Korentsides et al. 2024) and by the alignment of AI outputs with established cockpit workflows (Sadler et al. 2016). Moreover, studies of crew resource management indicate that insufficient transparency can trigger 'automation bias', where operators over-rely on, or conversely ignore, AI suggestions (Wickens et al. 2020). Thus, an important prerequisite for introducing AI to aircraft is the development of trust in the system. It must be ensured that pilots understand how the AI operates and know its capabilities and limitations in order to build adequate trust (Lee and See 2004).

Trust development

Trust in automation

Hoff and Bashir (2015) conceptualise trust in automation as a hierarchical construct composed of three interacting layers that together determine an operator's willingness to rely on a system. (1) Dispositional trust is a relatively stable, trait-like tendency of an individual to trust technology in general. Factors that influence dispositional trust include culture, age, gender and personality traits. (2) Situational trust is the initial trust, a judgement formed during the first exposure to a specific system, based on observable system attributes and contextual cues. Factors that influence situational trust include, e.g. type of system, system complexity, task difficulty, workload, perceived risks and benefits, and own expertise. (3) Learned trust is the evolving trust state that results from ongoing interaction, feedback, and performance monitoring – this layer is differentiated into initial learned trust prior to interaction, e.g. through pre-existing knowledge, such as an understanding of the system, and dynamic learned trust during interaction, such as by observing the system's performance, experiencing its design and usability and forming intentions to use it (Hoff and Bashir 2015). Thus, following this model, trust in automation is not a unidimensional, stable construct but is shaped by individual characteristics of the operator and the situation, and is formed over time through interactions with the system. Because of its universal nature, the model can be extended from automation in general to AI-based systems (Alarcon and Capiola 2025; Mehrotra et al. 2024; Okamura and Yamada 2020; Vorm and Combs 2022).

Increasing trust via transparency

The characteristics and capabilities of a system contribute significantly to the operator's trust in it. When operators have a clear understanding of the system through pre-existing knowledge, this will inform their initial and dynamic learned trust (Hoff and Bashir 2015). In established trust models, the understanding of a system can be informed by dimensions of Ability/Performance, Integrity/Process and Purpose/Benevolence (Lee and See 2004; Mayer, Davis, and Schoorman 1995). In short, the dimensions describe what the system does to help the operator achieve their goals, how the system works and whether the operator perceives the system as beneficial for their goals (see Appendix Table A1; Lee and See 2004; Mayer, Davis, and Schoorman 1995). Those dimensions have often been collapsed as 'transparency' information.

A growing consensus holds that transparency – the provision of understandable information about the inner workings, data sources, and decision logic of an AI system – can foster trust (Hoff and Bashir 2015; Lee and See 2004; Zerilli, Bhatt, and Weller 2022). Transparency may influence the aforementioned factors of trust development positively by providing the operator with a better basis for evaluating the system, its capabilities and its associated risks when using it, thus reducing uncertainty and the 'black box' problem typically associated with non-deterministic systems. Recent experimental work demonstrates that exposing operators to algorithmic transparency by means of explanations of algorithms, confidence scores, or visualisations of the decision process of AI-based systems leads to increases in trust and acceptance (Gegoff et al. 2024; Kizilcec 2016; Tatasciore, Strickland, and Loft 2024). In short, algorithmic transparency seems to be a key factor in operators' development of trust in AI-based systems.

Current study

In the current study, we examined the effect of the algorithmic transparency of an AI-based cockpit assistant on learned trust in pilots across two time points while controlling for important factors of situational trust (see Figure 1). We used the concepts of purpose/benevolence, integrity/process and ability/performance from the trust models (Lee and See 2004; Mayer, Davis, and Schoorman 1995) to establish the basis for algorithmic transparency.

We hypothesised that providing transparency via algorithmic explanations about the system's process and performance prior to interaction will increase trust ratings compared with those for explanations provided after interaction (no pre-disclosure). To investigate the effect of algorithmic transparency, we introduced a new assistant in the cockpit to subject matter experts (commercial airline pilots) and measured situational and learned trust during a tutorial session and a real flight.

We predicted trust indicators using the transparency condition and analysed the data to answer the following research questions:

RQ1) Do participants receiving information about the system early on (pre-interaction disclosure) have fewer demands for explainability after the tutorial than participants receiving the same information at the end of the experiment (manipulation check)?

RQ2) Does the transparency condition predict learned trust during the tutorial and the experimental flight?

RQ3) Can changes in trust also be predicted by the transparency condition (exploratory analysis)?

This work contributes to the current discussions on trust by extending the transparency–trust literature into the domain-specific ergonomics of AI-assisted aviation and by quantifying how trust evolves under differing explanatory modalities across extended mission timelines.

Methods

Participants

The study was conducted with 21 participants (21 male): all were pilots with an Airbus type rating. Their flight experience was variable. Average age was 40 years (range: 26–68), and average flight hours were 7657 (range: 1700–31500). Six pilots were captains and 15 pilots were first officers; refer Appendix Table A4 for an overview of the characteristics across groups. Ten pilots mainly conduct short-haul and 11 pilots mainly long-haul flights. All had normal hearing and normal or corrected vision conditions. The study was approved by the local ethics committee (number 32/25, Research Ethics Office of German Aerospace Center). All participants gave written informed consent. They received €40/h as compensation. The study is pre-registered: https://doi.org/10.17605/OSF.IO/KZTM3.

Design

The design of the study was a 2 (**transparency**: early and late) × 3 (**feature**: taxi, cruise, approach) × 2 (study **phase**: tutorial, flight) mixed design, with one between-subjects factor (transparency) and two within-subject factors (study phase and situation). Pilots conducted two flights during the experiment, one flight with and one without the assistant; the latter was not analysed for this study.

Each pilot participated alone, in an approximation of a single-pilot setting. Pilots were randomly assigned to either the early or the late **transparency** condition. In both conditions, pilots were verbally introduced to the aim and the scope of the assistant **features**. This introduction approximated the concepts of purpose/benevolence in trust models (Lee and See 2004; Mayer, Davis, and Schoorman 1995). In the early transparency condition, pilots were introduced to the internal details of the features before the tutorial – they were informed about the AI algorithms used, specific functionality and dependencies of the features from a technical



Figure 1. Indicators of the concepts in the framework of Hoff and Bashir (2015) that were used in the current study.

perspective, and accuracy rate. This information served as the concepts of integrity/process and ability/performance in trust models (Lee and See 2004; Mayer, Davis, and Schoorman 1995). In the late transparency condition, pilots were informed about these internal details only in the debriefing after the experiment. See Table A2 in the Appendix for an overview of how the features were introduced in each transparency condition.

Each pilot worked with the assistant during two study phases of the experiment. In the first study phase, the tutorial phase, they received a guided tour of the assistant with its three features and their functionalities and learned how to use them in one concrete situation. The tutorial consisted of three segments that did not belong to a coherent flight but were chosen to demonstrate the functionalities of the assistant during different flight phases. In the second study phase, the main experimental flight phase, the pilots conducted two flights, one with and one without the assistant. Two similar flights were designed (in the following, called 'blue' and 'yellow') and each pilot administered both flights. The order of the flights (blue vs. yellow, with vs. without assistant) was counterbalanced across participants. In this article, only the flights with the assistant are analysed, with half of the pilots flying the blue flight and half of the pilots flying the yellow flight.

Both flights departed from Madrid-Barajas Airport (LEMD) and had their destination at Frankfurt am Main Airport (EDDF). For 'blue' and 'yellow' flights, different technical problems before and during flight, different taxi routes, different weather events and different flight routes, resulting in different standard arrival routes (STARs) and arrival runways (RWY). See Appendix Table A3 for an overview of the differences between the flights. Similar to the tutorial, the main experimental flights were not simulated as a whole, but three segments were chosen that correspond to the three features of the assistant in a fixed order:

- Taxi: In this situation, the pilot needs to taxi the aircraft from the stand to the runway and take off. Here, the feature assists by transcribing air traffic controllers' or apron controllers' instructions and visualising them on a moving map.
- Cruise: In this situation, a weather condition will cross the flight route of the aircraft, and a diversion decision has to be executed. Here, the feature assists by visualising the location of the weather event and proposing a possible alternative diversion route.
- Approach and landing preparation (in the following 'Approach'): Approximately one hour before landing, the pilot checks if they want to update the flight plan for arrival. Here, the feature assists by predicting a likely STAR and runway for landing.

Procedure and questionnaires

The study was conducted via a flat-panel cockpit simulator for the Airbus A321 using X-Plane 11. The cockpit displays and assistant interfaces are visualised on touch screens, while sidesticks, pedals and the glareshield with the flight control unit (FCU) are hardware components. The assistant was displayed next to the ECAM display, embedded into a panel of the simulator. All participants were seated in the left seat of the flight simulator (Captain's position), independent of their current working position in real-life.

Pilots went through a short familiarisation session with the simulator and an unfamiliar single-pilot setting, and afterwards were introduced to the assistant. The experimenter gave them a step-by-step tutorial on where which information could be found and how the functionalities worked and how decisions could be made. Afterwards, they could try out the functionalities in the tutorial flights. After a break, the experimental flights took place. Participants were instructed to think aloud during the study and to assume that the duties of the second cockpit member, like reading out checklists, were duly performed.

We measured specific situational and learned trust indicators within the framework of Hoff and Bashir (2015); see Figure 1.

Covariates: situational trust

All situational trust variables were assessed prior to any information on the assistants after signing the declaration of consent.

For *subject matter expertise*, participants provided their cumulative number of flight hours as **flight hours**.

For *perceived risks*, we formulated items for the three features by implementing the Threats of Artificial Intelligence Scale (Kieslich, Lünich, and Marcinkowski 2021): two recognition risks for Feature Taxi and Takeoff, two recognition and three recommendation risks for Feature Cruise, and two prediction risks for Feature Approach. The scale is rated on a five-point Likert scale from non-threatening (1) to very threatening (5). We summed all answers to the items separated by feature as **risks**.

For *perceived benefits*, we used the AIAS-4 scale (Grassini 2023). The scale consists of four items and was rated on a ten-point Likert scale between Not at all (1) and Completely Agree (10). We summed all answers to

the items as **attitudes**. We further used the dimensions of Attainment and Utility Value of the Questionnaire of Artificial Intelligence Use Motives (Yurt and Kasarci 2024). The scales consist of four items in each dimension and were rated on a five-point Likert scale from completely false (1) to completely true (5). We summed all answers to the items as **attainment** and **utility**.

Dependent variables: manipulation check

To prove that the manipulation of transparency was successful, we assessed the dimensions of perceived own expertise and demand for explainability (Weber, Hußmann, and Eiband 2021) once after the tutorial phase was separated for each feature. The scales consist of two and four items, respectively, and were rated on a five-point Likert scale from strongly disagree (1) to strongly agree (5).

Dependent variables: learned trust

We collected predictability and reliability for all study phases, but ease-of-use, intention-of-use and communication style only after the flight due to feasibility reasons during data collection.

For **predictability** and **reliability**, participants filled in the Trust in Automation Scale (Körber 2019) after the tutorial and after the experimental flight, separated by each Feature. The scale consists of several dimensions; for the purpose of this study, we collected predictability and reliability. The dimensions were rated on a five-point Likert scale from strongly disagree (1) to strongly agree (5).

For **ease-of-use**, participants filled in the System Usability Scale (Brooke 1996) after the experimental flight separated for each Feature. The questionnaire consists of ten questions on a five-point Likert scale, from strongly disagree (1) to strongly agree (5).

For **intention-to-use**, participants also rated their intention to use the corresponding Feature on a scale from strongly disagree (1) to strongly agree (5) after each Feature flight.

For **communication style**, participants rated the SHAPE teaming questionnaire (Dehn 2008; Jipp and Papenfuss 2011). We selected five items that were appropriate in the context of the used assistant (1, 4, 10, 11 and 15). Pilots rated them on a five-point Likert scale from strongly disagree (1) to strongly agree (5).

Deviations from pre-registration

We additionally preregistered to measure further constructs of trust, namely safety and the SATI questionnaire, but did not use them in this study. The SATI had several overlaps with the reliability and predictability dimensions of the TiA scale and overlaps with usability items in the System Usability Scale; therefore, we decided not to reduplicate items. The answers to the safety question were a mixture of comments about the study design, study setup, technical problems of the simulator or the simulation, and relations to the Feature; we thus decided to omit this question as measurement error was too high.

Analyses

In a covariate analysis, we investigated which situational trust indicators were related to learned trust to select those indicators as covariates in the main transparency analyses. Attainment, utility and attitudes were non-significant, whereas flight hours and risks predicted learned trust. We therefore included flight hours and risks in the following main analyses.

We investigated three research questions with the help of mixed-effect cumulative link models (package ordinal) in R (R Core Team 2025). The first question investigated whether the manipulation was successful and whether participants had fewer demands for explainability after the tutorial. The second question investigated whether the transparency condition predicted learned trust during the tutorial and during the experimental flight. The third question investigated whether the changes in trust can also be predicted by the transparency condition.

For questions 1 and 2, we predicted the ordinal answers of the corresponding trust questionnaires, and for question 3, we predicted the difference between flight and tutorial. In all analyses, we included as predictors the transparency condition in interaction with dimension (demand for explainability, perceived own knowledge, predictability, reliability, ease-of-use, intention-of-use, communication style) and feature, with the covariates flight hours and risks as well as a random intercept for participant ID. For question 3, we further added the answer given for the tutorial for each respective question item. For all questions, we initially examined whether Feature had interactions with the Transparency condition; if yes, we report this less parsimonious model; if no, we report it without interaction. Sensitivity analyses were conducted using nonparametric bootstrap resampling (1000 replications). For all effects, bootstrap estimates, bias, and standard errors were calculated and compared with the corresponding estimates from the original model to assess the robustness of the results.

For all models, we report the corresponding ANOVA Type III results (package car), the cumulative link model

(package ordinal) estimates with bootstrap results (package boot) and post hoc Tukey tests (package emmeans).

Deviations from pre-registration

We initially planned to analyse the research questions with an interaction model between study phase and transparency condition for each scale. Due to the small sample size, we were not able to run this interaction model. We thus split the models according to study phase (tutorial, flight, change between tutorial and flight) and reduced the number of anticipated models by collapsing the scales.

Results

We report the ANOVA results of all models in Appendix Table A5 and all original cumulative link model results in Appendix Table A6.

Was the manipulation of transparency successful?

The cumulative link model for the manipulation check (Cond. R^2 = 0.425, Marg. R^2 = 0.312) showed main effects for transparency (higher scores in early transparency compared to late transparency), feature (higher scores for Feature Taxi compared to Approach), and risk for Approach (higher risk is associated with lower scores), as well as an interaction between transparency and dimension (perceived own expertise is more strongly influenced than demand for explainability); see Appendix Tables A5 and A6.

Is transparency associated with higher trust?

The cumulative link model for the tutorial flight (Cond. R^2 = 0.429, Marg. R^2 = 0.302) showed main effects for transparency (higher scores in early transparency compared to late transparency), dimension (predictability showed higher scores than reliability), feature (higher trust for Feature Cruise compared to Taxi and to Approach) and risks for Cruise and for Approach (higher risk is associated with lower scores); see Appendix Tables A5 and A6.

The cumulative link model for the experimental flight showed main effects for transparency (higher scores in early transparency compared to late transparency), dimension (ease-of-use showed higher scores than reliability and communication style; predictability and communication style showed higher scores than reliability), feature (higher trust for Feature Cruise compared to Taxi and to Approach), flight hours (more flight hours are associated with higher trust scores), and risks for Cruise; see Appendix Table A2. Furthermore, interactions between transparency and dimension (the transparency effect was strongest for ease-of-use) and dimension and feature were observed; see Figure 2.

Is transparency associated with a change in trust during interaction?

The cumulative link model for the difference between experimental flight and tutorial (Cond. R^2 = 0.511, Marg. R^2 = 0.462) showed main effects for tutorial score (lower tutorial scores are associated with higher change in trust), transparency (higher increase in early transparency compared to late transparency), dimension (predictability showed higher increases than reliability), feature (higher trust for Features Taxi and Cruise compared to Approach), flight hours (more flight hours are associated with higher trust scores) and risks for Cruise and for Approach (higher risk is associated with lower scores); see Appendix Tables A5 and A6 and Figure 3.

Discussion

In this study, transparency was operationalised as an experimental manipulation by informing pilots of the internal details of the assistant features either early (before the tutorial) or late (after the experiment during the debriefing). The two transparency conditions thus differed in terms of the timing at which participants were provided with information on the underlying AI algorithms, their functional dependencies, and already known performance metrics (accuracy, integrity, and performance). The results showed that early transparency increased trust following the tutorial, following the actual simulated flight, and that the increase in trust from the tutorial to the simulated flight is significantly greater for pilots exposed to early transparency than for those in late transparency conditions. The bootstrap sensitivity analyses indicated high stability of the parameter estimates. This supports the hypothesis that early insight into how the assistant works strengthens trust not only immediately after instruction but also throughout the entire duration of the mission (Hoff and Bashir 2015).

Acknowledgements

We would like to thank the pilots for their participation in the study and Viola Kershi for assisting in data collection and data management.

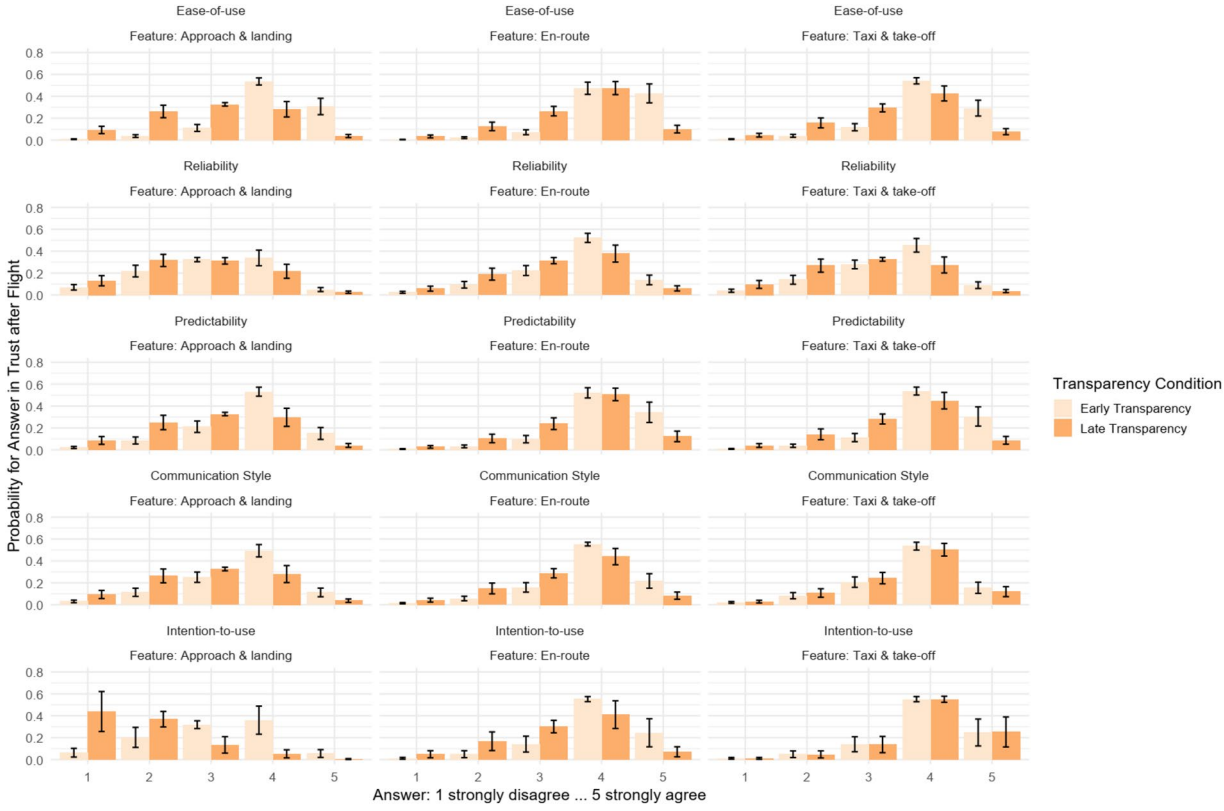


Figure 2. Predicted probabilities for the experimental flight in the two transparency versions separated by dimension and feature.

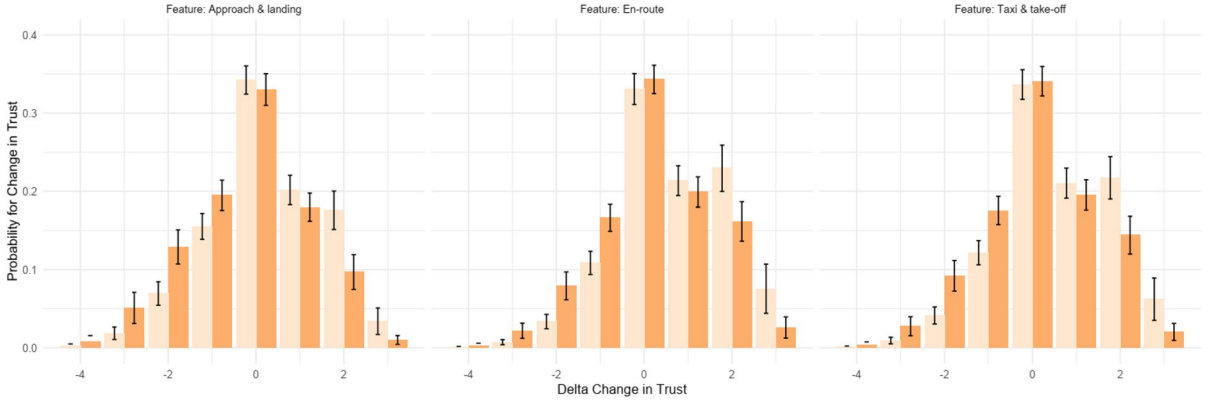


Figure 3. Predicted probabilities for change in trust in both transparency versions separated by feature. A negative score refers to lower trust in the experimental flight compared to the tutorial flight and a positive score reflects higher trust.

Transparency in trust frameworks

The transparency manipulation provided pilots with accuracy, integrity and performance details. For our purposes, this information serves as the dimensions of Purpose/Benevolence and Ability/Integrity (Lee and See 2004; Mayer, Davis, and Schoorman 1995). Through the early disclosure of technical details (e.g. the machine learning model used, the training dataset and the proven success rates), pilots are immediately provided with a basis for assessing the system's ability. The empirically observed increase in trust following the tutorial suggests that this information might be effectively integrated into the pilots' mental models and enhances the perceived competence of the system. Early information about process integrity (e.g. which sensors/information influence which decisions, which validation mechanisms are implemented) might act as a sign of transparency and reduce the risk of 'black-box' scepticism. Finally, the knowledge that the system might serve the mission's safety objective (purpose/benevolence component; Lee and See 2004) significantly increases trust following the tutorial.

The significantly greater increase in trust from the tutorial to the flight phase among pilots in the early transparency condition might be interpreted as trust consolidation. People adjust their trust when observed events match previously set expectations. Pilots who are informed early on can develop expectations regarding accuracy, error rates and decision-making logic prior to interaction. During flight, these expectations can be confirmed by the behaviour actually observed, and the resulting positive confirmation effect increases and reinforces trust (Lee and See 2004). In contrast, participants in the late

transparency group must first develop an intuitive, often vague picture of the system; only retrospectively might this picture be corrected by technical performance.

The pattern of results may additionally be shaped by cognitive resources (Loock 2025; Risko and Gilbert 2016; Sweller 2020). Early transparency can reduce intrinsic and extrinsic cognitive load because pilots can develop a structured knowledge framework during disclosure that facilitates the processing of new information. During the tutorial, complex system principles can be internalised; during the simulated flight, the pilot only needs to apply this knowledge rather than learn it. The cognitive resources freed up can be used to assess system performance, which might further increase confidence.

Further trust predictors: System complexity, risks, subject-matter expertise

The present finding suggests that trust depends on specific system features, particularly their functionality within distinct flight phases. This aligns not only with the multi-dimensional trust framework by Hoff and Bashir (2015) but also with research showing that reliance on decision aids depends on feature-level characteristics, such as perceived accuracy and accountability, rather than on the system as a whole (Skitka, Mosier, and Burdick 2000). Moreover, automation use and trust vary across flight phases, with higher reliance in stable, low-workload conditions such as cruise and more selective or cautious use during dynamic phases such as approach and landing (Billings 2018; Dismukes, Berman, and Loukopoulos 2017; Wiener 1989; Young and Stanton 2023). These differences are often attributed to variations in workload, predictability, and mode complexity, indicating that trust emerges from the interaction between feature capability and situational demands, rather than reflecting a stable system-level attitude. This shows that trust in automation is multi-faceted and function-specific rather than global and should be assessed as such.

The present findings further indicate that trust increases as perceived risk decreases. This relationship is well supported in the literature, as risk perception and system reliability are central determinants of trust, particularly in high-stakes environments such as aviation (Lee and See 2004; Mayer, Davis, and Schoorman 1995). Systems perceived as less risky are more likely to be trusted and used, whereas high perceived risk can lead to distrust or disuse (Chancey et al. 2025). Thus, even highly reliable systems may be distrusted if perceived risks remain high. In the current study, we explicitly investigated pilots, a sample who is more likely to focus on the safety aspects of operations and systems than the general population due to specialised training. Even though their attitude towards automation and AI might be more negative than in the general population, we detected an association between their system perception and trust building. This effect may pose a problem, especially for the aviation domain, and further highlights the necessity to develop reliable and transparent assistants and invest in trust-building early on.

A further result suggests that trust increases with subject-matter expertise (i.e., flight hours). While this is intuitively plausible, prior research indicates a more differentiated relationship. Experience improves operators' understanding of system capabilities (Parasuraman and Riley 1997), which can support more informed trust. Accordingly, greater flight experience may foster more stable trust judgements. However, empirical work also demonstrates that expertise does not uniformly increase trust. Instead, experienced operators tend to develop more appropriate and often more critical trust. Appropriate trust reflects a match between trust and system performance rather than maximisation (Lee and See 2004). In line with this, experienced pilots might be better able to detect system limitations and adjust their reliance accordingly. However, flight experience is also highly associated with age (in our sample $r(20) = 0.88$, $p < 0.001$), a further predictor of trust (Olson, Fisk, and Rogers 2009; Schaefer et al. 2016). Future studies are needed to shed more light on disentangling the association between experience, age and trust.

Future directions

In the present research, transparency was only manipulated by giving information early or late in the process. Future research should also address exactly what kind of information is needed for building trust, such as how the system works exactly (e.g. used algorithms, training data) or what its strengths and limitations are (e.g. prediction accuracy, edge cases). Furthermore, separation of process transparency and performance transparency would be helpful in understanding the mechanisms of the trust-building processes in greater detail. It should also be examined whether early transparency information reduces automation bias without sacrificing appropriate reliance, an essential balance for safety-critical operations (Endsley 2017). Moreover, the change in information processing requirements and downstream functional task requirements associated with having to assimilate information provided by the assistant systems might be an additional relevant indicator (Fan et al. 2026; Hermanns and Teubner 2026). Furthermore, future research can include the actual trustworthiness of the system to shed light on the interplay between the actual performance of an assistance system and the perceived trustworthiness of the user (Schlicker et al. 2025), as well as the reliance on the system as a behavioural indicator of trust (Hoff and Bashir 2015).

Practical implications

- **Provide explanations early** – Giving pilots information early on seems to positively influence trust in the system. It might be beneficial if designers of cockpit AI deliver concise, flight-phase-specific explanations before pilots interact with the system (e.g., during initial aircraft-specific flight training, via pre-flight briefings or on-screen 'why' pop-ups). Training and briefing materials could potentially be improved by clearly explaining the algorithmic principles, error rates and validation procedures before pilots have their first practical experience with the system. The density of information should be adapted to the pilots' cognitive workload (Kizilcec 2016).
- **Leverage pilot expertise** – Experienced pilots seem to respond positively to transparency; training programmes on new systems could emphasise how explanations complement existing mental models,

thereby hopefully fostering higher trust among novices. Experienced pilots already possess well-structured, domain-specific knowledge frameworks into which explanatory information can be seamlessly integrated, for example, through scenario-based walkthroughs with annotated decision trees or through layered transparency panels that display additional technical details when required. For trainee pilots, early, structured explanations could not only raise the initial level of trust but also accelerate the learning curve, as they provide a consistent reference framework for subsequent independent action.

- **Mitigate perceived risk** – Perceived risk likely decreases trust. A possible mitigation could be to include transparent risk indicators (e.g., confidence intervals, uncertainty bounds) together with explanations to counteract the negative effect of perceived threats (Dwivedi et al. 2023; Holzinger, Carrington, and Müller 2020). If AI is used, it should be explainable AI, i.e. a suite of machine learning techniques that enable human users to understand, appropriately trust, and produce more explainable models (Hernandez, Ayo, and Panagiotakopoulos 2021; Hoffman et al. 2023).

Limitations

Although a flight simulator offers high fidelity, it cannot fully replicate the stressors and consequences of real-world aviation, potentially inflating trust levels. Trust was measured only across one tutorial and one experimental flight. Longitudinal designs are needed to assess the durability of transparency effects over multiple missions to fully investigate the trust evolution (Hoff and Bashir 2015; Muir 1994).

The study collapsed some flight phases into a single 'feature' label. Within-phase variability (e.g., different weather conditions, different complexities of airport layouts and aircraft failures) may moderate transparency effects and warrant finer-grained analysis investigating more situational factors. In addition, because the features were designed to assist during different flight phases, they had different assistance functionalities, differed in the usability of the interfaces, and were based on different AI or machine learning algorithms; they also differed in operational demands, and we used a fixed presentation order. Thus, further studies are needed to generalise the results.

Finally, in the present study, an artificial single-pilot setting was chosen to maximise the sample size and minimise crew interaction effects, which may have impacted the development of trust in the systems. Since there was no option to cross-check options and decisions with a colleague (Hörmann 1994), the pilots may have relied more on the assistant as an additional source of information (Hollan, Hutchins, and Kirsh 2000; Hutchins 2006; Loock 2025), thus artificially increasing trust. Alternatively, this setting may have negatively influenced trust because the sole responsibility for each decision lay with the single pilot. Research shows that the pilot flying perceives more risks and is more emotionally involved than the pilot monitoring (Gao et al. 2025). This effect may even be more pronounced in a single-pilot setting, leading to more critical evaluation of feature behaviour. Future studies should therefore investigate differences between single-pilot and dual-crew settings as well as pilots' roles in the development of trust.

Conclusion

Early, phase-specific algorithmic transparency markedly increased pilots' situational and learned trust in the AI-driven assistants, as evidenced by higher trust ratings after the tutorial, after the flight, and a larger trust gain between the two phases. The effect was strongest for ease-of-use and was amplified for pilots with more flight hours, indicating that detailed explanations integrate readily into existing mental models. These findings suggest that pre-flight briefings containing layered, technical explanations should become a standard element of cockpit AI training to foster calibrated reliance. Future work should test adaptive, workload-aware transparency and extend the approach to dual-crew settings to confirm its robustness in operational aviation.

Author contributions

CRediT: **Anneke Hamann**: Conceptualization, Investigation, Methodology, Project administration, Resources, Writing – original draft; **Alexander Heise**: Methodology, Resources, Validation, Writing – review & editing; **Helge Lenz**: Methodology, Resources, Software, Validation, Writing – review & editing; **Anja Gampe**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Visualization, Writing – original draft.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Funding

JARVIS has received funding from the SESAR Joint Undertaking under the European Union's Horizon Europe research and innovation programme 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.

ORCID

Anneke Hamann http://orcid.org/0000-0001-6312-5116
Alexander Heise http://orcid.org/0009-0006-2380-7327
Helge Lenz http://orcid.org/0009-0002-9503-6403
Anja Gampe http://orcid.org/0000-0001-9812-9694

Data availability statement

The data that support the findings of this study are openly available in OSF at https://doi.org/10.17605/OSF.IO/KZTM3.

References

Alarcon, G. M., and A. Capiola. 2025. "Explicating the Trust Process for Effective Human Interaction with Artificial Intelligence and Machine Learning Systems". *Frontiers in Computer Science* 7: 1662185. doi:10.3389/fcomp.2025.1662185.

Billings, C. E. 1991. "Toward a Human-Centered Aircraft Automation Philosophy". *The International Journal of Aviation Psychology* 1 (4): 261–270. doi:10.1207/s15327108ijap0104_1.

Billings, C. E. 2018. *Aviation Automation: The Search for A Human-Centered Approach*. CRC Press.

Brooke, J. 1996. SUS - A Quick and Dirty Usability Scale. In *Usability Evaluation In Industry* (1st Edition). CRC Press.

Candrian, C., and A. Scherer. 2022. "Rise of the Machines: Delegating Decisions to Autonomous AI." *Computers in Human Behavior* 134: 107308. doi:10.1016/j.chb.2022.107308.

Carabantes, M. 2020. "Black-Box Artificial Intelligence: An Epistemological and Critical Analysis." *AI & Society*. 35 (2): 309–317. doi:10.1007/s00146-019-00888-w.

Chancey, E. T., M. S. Politowicz, K. M. Ballard, J. Unverricht, B. K. Buck, and S. Geuther. 2025. "Human-Automation Trust Development as a Function of Automation Exposure, Familiarity, and Perceived Risk: A High-Fidelity Remotely Operated Aircraft Simulation." *Journal of Cognitive Engineering and Decision Making* 19 (2): 200–222. doi:10.1177/15553434241296573.

Dehn, D. M. 2008. "Assessing the Impact of Automation on the Air Traffic Controller: The SHAPE Questionnaires." *Air Traffic Control Quarterly* 16 (2): 127–146. doi:10.2514/atcq.16.2.127.

Dismukes, R. K., B. A. Berman, and L. Loukopoulos. 2017. *The Limits of Expertise: Rethinking Pilot Error and the Causes of Airline Accidents*. Routledge. doi:10.4324/9781315238654.

Dwivedi, R., D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan. 2023. "Explainable AI (XAI): Core Ideas, Techniques, and Solutions." *ACM Computing Surveys* 55 (9): 1–33. doi:10.1145/3561048.

EASA 2023. "*Artificial Intelligence Roadmap 2.0: Human-Centric Approach to AI in Aviation* (Version 2.0)." European Union Aviation Safety Agency. https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-roadmap-20

Endsley, M. R. 2017. "From Here to Autonomy: Lessons Learned from Human–Automation Research." *Human Factors* 59 (1): 5–27. doi:10.1177/0018720816681350.

Fan, G., D. Liu, L. Pan, and R. Zhang. 2026. "When Help Hurts: Verification Load and Fatigue with AI Coding Assistants." *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, CHI '26*, 1–25. doi:10.1145/3772318.3791176.

Feary, M. 2018. "*A First Look at the Evolution of Flight Crew Requirements for Emerging Market Aircraft*." AIAA Aviation 2018, Atlanta, GA. https://ntrs.nasa.gov/citations/20180005720

Gajduskova, T., E. Jencova, A. Turac, D. Majorska, K. Hreskova, and G. Galgoci. 2025. "Regulatory Framework of EASA, European Strategies, and Practical Applications of Artificial Intelligence in Aviation." *2025 New Trends in Aviation Development (NTAD)*, 78–83. doi:10.1109/NTAD67887.2025.11302664.

Gao, S., L. Wang, N. Zhang, Y. Bai, Y. Tian, S. Chen, Y. Zhang, and H. Zhou. 2025. "Pilot Flying versus Pilot Monitoring: The Effect of Role Assignment on Pilots' Perceived Risk in Flight Simulation Experiment." *Risk Analysis: An Official Publication of the Society for Risk Analysis* 45 (9): 2759–2770. doi:10.1111/risa.70048.

Gegoff, I., M. Tatasciore, V. Bowden, J. McCarley, and S. Loft. 2024. "Transparent Automated Advice to Mitigate the Impact of Variation in Automation Reliability." *Human Factors* 66 (8): 2008–2024. doi:10.1177/00187208231196738.

Grassini, S. 2023. "Development and Validation of the AI Attitude Scale (AIAS-4): A Brief Measure of General Attitude toward Artificial Intelligence." *Frontiers in Psychology* 14: 1191628. doi:10.3389/fpsyg.2023.1191628.

Halawi, L., M. Miller, and S. Holley. 2024. "Artificial Intelligence in Aviation: A Path Analysis." *Journal of Aviation/Aerospace Education & Research* 33 (4) doi:10.58940/2329-258X.2061.

Hermanns, L., and T. Teubner. 2026. "Under Pressure: How Time Constraints, Task Complexity, and AI Reliability Shape human-AI Interaction." *Behaviour & Information Technology* 45 (11): 2597–2621. doi:10.1080/0144929X.2025.2587732.

Hernandez, C. S., S. Ayo, and D. Panagiotakopoulos. 2021. "An Explainable Artificial Intelligence (xAI) Framework for Improving Trust in Automated ATM Tools." *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)*, 1–10. doi:10.1109/DASC52595.2021.9594341.

Hoff, K. A., and M. Bashir. 2015. "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust." *Human Factors* 57 (3): 407–434. doi:10.1177/0018720814547570.

Hoffman, R. R., S. T. Mueller, G. Klein, and J. Litman. 2023. "Measures for Explainable AI: Explanation Goodness, User Satisfaction, Mental Models, Curiosity, Trust, and human-AI Performance." *Frontiers in Computer Science* 5 doi:10.3389/fcomp.2023.1096257.

Holford, W. D. 2022. "An Ethical Inquiry of the Effect of Cockpit Automation on the Responsibilities of Airline Pilots: Dissonance or Meaningful Control?" *Journal of Business Ethics* 176 (1): 141–157. doi:10.1007/s10551-020-04640-z.

Hollan, J., E. Hutchins, and D. Kirsh. 2000. "Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research." *ACM Transactions on Computer-Human Interaction* 7 (2): 174–196. doi:10.1145/353485.353487.

Holzinger, A., A. Carrington, and H. Müller. 2020. "Measuring the Quality of Explanations: The System Causability Scale (SCS). Comparing Human and Machine Explanations." *KI - Künstliche Intelligenz* 34 (2): 193–198. doi:10.1007/s13218-020-00636-z.

Hörmann, H. J. 1994. "*For-DEC - A Prescriptive Model for Aeronautical Decision Making*." 21. *WEAAP-Conference*, Dublin, 28. https://elib.dlr.de/27044/.

Hutchins, E. 2006. "The Distributed Cognition Perspective on Human Interaction." In *Roots of Human Sociality*. Routledge.

Jipp, M., and A. Papenfuss. 2011. "Reliability and Validity of the SHAPE Teamwork Questionnaire." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 55 (1): 115–119. doi:10.1177/1071181311551024.

Kieslich, K., M. Lünich, and F. Marcinkowski. 2021. "The Threats of Artificial Intelligence Scale (TAI)." *International Journal of Social Robotics* 13 (7): 1563–1577. doi:10.1007/s12369-020-00734-w.

Kizilcec, R. F. 2016. "How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface." *Proceedings*

of the 2016 CHI Conference on Human Factors in Computing Systems, 2390–2395. doi:10.1145/2858036.2858402.

Körber, M. 2019. "Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation." In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Hrsg.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (Bd. 823, S13–30 Springer International Publishing. doi:10.1007/978-3-319-96074-6_2.

Korentsides, J., J. Keebler, C. Fausett, S. Patel, and E. Lazzara. 2024. "Human-AI Teams in Aviation: Considerations from Human Factors and Team Science." *Journal of Aviation/Aerospace Education & Research* 33 (4). doi:10.58940/2329-258X.2046.

Lee, J. D., and K. A. See. 2004. "Trust in Automation: Designing for Appropriate Reliance." *Human Factors* 46 (1): 50–80. doi:10.1518/hfes.46.1.50_30392.

Loock, L. 2025. 'The Extracted Mind.' *Synthese* 205 (3): 1–23. doi:10.1007/s11229-025-04962-3.

Mayer, R. C., J. H. Davis, and F. D. Schoorman. 1995. "An Integrative Model of Organizational Trust." *The Academy of Management Review* 20 (3): 709–734. doi:10.2307/258792.

Mehrotra, S., C. Degachi, O. Vereschak, C. M. Jonker, and M. L. Tielman. 2024. "A Systematic Review on Fostering Appropriate Trust in Human-AI Interaction: Trends, Opportunities and Challenges." *ACM Journal on Responsible Computing* 1 (4): 1–45. doi:10.1145/3696449.

Muir, B. M. M. 1994. "Trust in Automation: Part I. Theoretical Issues in the Study of Trust and Human Intervention in Automated Systems." *Ergonomics* 37 (11): 1905–1922. doi:10.1080/00140139408964957.

Okamura, K., and S. Yamada. 2020. "Adaptive Trust Calibration for human-AI Collaboration." *PloS One* 15 (2): e0229132. doi:10.1371/journal.pone.0229132.

Olson, K. E., A. D. Fisk, and W. A. Rogers. 2009. "Collaborative Automated Systems: Older Adults' Mental Model Acquisition and Trust in Automation." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 53 (22): 1704–1708. doi:10.1177/154193120905302210.

Parasuraman, R., and V. Riley. 1997. "Humans and Automation: Use, Misuse, Disuse, Abuse." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39 (2): 230–253. doi:10.1518/001872097778543886.

R Core Team 2025. *R: A Language and Environment for Statistical Computing*. *R Foundation for Statistical Computing* [Software]. https://www.R-project.org.

Risko, E. F., and S. J. Gilbert. 2016. "Cognitive Offloading." *Trends in Cognitive Sciences* 20 (9): 676–688. doi:10.1016/j.tics.2016.07.002.

Sadler, G., H. Battiste, N. Ho, L. Hoffmann, W. Johnson, R. Shively, J. Lyons, and D. Smith. 2016. "Effects of Transparency on Pilot Trust and Agreement in the Autonomous Constrained Flight Planner." *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, 1–9. doi:10.1109/DASC.2016.7777998.

Schaefer, K. E., J. Y. C. Chen, J. L. Szalma, and P. A. Hancock. 2016. "A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems." *Human Factors* 58 (3): 377–400. doi:10.1177/0018720816634228.

Schlicker, N., K. Baum, A. Uhde, S. Sterz, M. C. Hirsch, and M. Langer. 2025. "How Do We Assess the Trustworthiness of AI? Introducing the Trustworthiness Assessment Model (TrAM)." *Computers in Human Behavior* 170: 108671. doi:10.1016/j.chb.2025.108671.

Sheridan, T. B. 1999. "Human Supervisory Control of Aircraft, Rail and Highway Vehicles." *Transactions of the Institute of Measurement and Control* 21 (4-5): 191–201. doi:10.1177/014233129902100407.

Skitka, L. J., K. Mosier, and M. D. Burdick. 2000. "Accountability and Automation Bias." *International Journal of Human-Computer Studies* 52 (4): 701–717. doi:10.1006/ijhc.1999.0349.

Sweller, J. 2020. "Cognitive Load Theory and Educational Technology." *Educational Technology Research and Development* 68 (1): 1–16. doi:10.1007/s11423-019-09701-3.

Tatasciore, M., L. Strickland, and S. Loft. 2024. "Transparency Improves the Accuracy of Automation Use, but Automation Confidence Information Does Not." *Cognitive Research: principles and Implications* 9 (1): 67. doi:10.1186/s41235-024-00599-x.

Taylor, A. K., and T. S. Cotter. 2020. "Pilots' Role in the Critical Infrastructure of Aviation." In I. L. Nunes (Hrsg.), *Advances in Human Factors and Systems Interaction* (pp 349–360). Springer International Publishing. doi:10.1007/978-3-030-20040-4_32.

Vorm, E. S., and D. J. Y. Combs. 2022. "Integrating Transparency, Trust, and Acceptance: The Intelligent Systems Technology Acceptance Model (ISTAM)." *International Journal of Human–Computer Interaction* 38 (18-20): 1828–1845. doi:10.1080/10447318.2022.2070107.

Weber, T., H. Hußmann, and M. Eiband. 2021. "Quantifying the Demand for Explainability." In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, & K. Inkpen (Hrsg.), *Human-Computer Interaction – INTERACT 2021* (Bd. 12933, S. 652–661). Springer International Publishing. doi:10.1007/978-3-030-85616-8_38.

Wickens, C. D., and F. Dehais. 2019. 'Expertise in Aviation." In P. Ward, J. M. Schraagen, J. Gore, & E. M. Roth (Hrsg.), *The Oxford Handbook of Expertise*. Oxford University Press.

Wickens, C. D., N. J. Fitzgerald, B. A. Clegg, C. A. P. Smith, D. Orth, and K. Kincaid. 2020. "Decision Aiding for Nautical Collision Avoidance: Trust, Dependence, and Implicit Understanding of the Decision Algorithm." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 64 (1): 1950–1954. doi:10.1177/1071181320641470.

Wiener, E. L. 1989. Human factors of advanced technology (glass cockpit) transport aircraft (NAS 1.26:177528). https://ntrs.nasa.gov/citations/19890016609

Young, M. S., and N. A. Stanton. 2023. *Driving Automation: A Human Factors Perspective*. CRC Press. doi:10.1201/9781003374084.

Yurt, E., and I. Kasarci. 2024. "A Questionnaire of Artificial Intelligence Use Motives: A Contribution to Investigating the Connection between AI and Motivation." *International Journal of Technology in Education* 7 (2): 308–325. doi:10.46328/ijte.725.

Zerilli, J., U. Bhatt, and A. Weller. 2022. 'How Transparency Modulates Trust in Artificial Intelligence." *Patterns (New York, N.Y.)* 3 (4): 100455. doi:10.1016/j.patter.2022.100455.

Appendix

Table A1. Dimensions informing system understanding according to Lee and See (2004) and Mayer, Davis, and Schoorman (1995).

System understanding	Dimension	Description
Does the system help me?	Ability	Trust based on an assessment of the other party's technical competence
	Performance	The extent to which the system operates in a way that helps the operator achieve their goals
How does the system work?	Integrity	The assessment that the counterpart acts honestly, reliably and in compliance with rules
	Process	The evaluation if the system and its algorithms are appropriate for the situation
Do my goals and the system capabilities align?	Benevolence	The extent to which the intents and motivations of the system and the operator align
	Purpose	The knowledge about the intended application of the system

Table A2. Transparency information given for the features of the assistant.

Condition	Feature Taxi and takeoff	Feature En-route	Feature Approach and landing
Information about aims and scope given during the tutorial in **both transparency conditions**	**Benevolence/Purpose**: This feature is designed to help during the entire flight. It supports pilots with ATC instructions and clearances. It listens to ATC communication directed to and from the aircraft. The feature transcribes the speech into text, analyses the textual context and displays it to the pilot on a screen (textually and sometimes graphically).	**Benevolence/Purpose**: This feature can be used in climb, cruise and descent. It supports the pilot's context awareness and decision-making. To do this, it constantly monitors the flight context both internally in terms of aircraft status and performance and externally via NOTAMs, METARs, SIGMETs and PIREPs for unexpected events that could influence the current flight strategy. If an event or combination of events is considered risky for the current flight strategy, the system notifies the pilots and supports them in making decisions on how to deal with it.	**Benevolence/Purpose**: This feature is intended to help during the approach and landing phase. It supports pilots in preparing for arrival and landing. It generates predictions about the standard arrival route and the runway that could be assigned to the flight. These predictions can be generated before and during the cruise phase, giving the pilot more time to prepare.
Information about internal details is given **before the tutorial in the early transparency** condition but **during debriefing in the late transparency** condition	**Integrity/Process**: This feature is implemented with algorithms and AI components. The AI components are embedded in the CYRUS language processing engine. The speech processing engine transcribes speech into text and labels sentences based on a grammar given as input. Speech recognition uses hybrid models (in this case, Hidden Markov Models with a Deep Neural Network) for acoustic analysis. An algorithm was then developed based on the keyword spotting function to carry out the syntactic analysis of the ATC and pilot sentences for ADA. **Ability/performance**: The CYRUS model was mainly trained using speech-driven ATC training simulators in 20 different countries. Data from both men and women are included, with low background noise (mainly background noise from other air traffic controllers working at nearby stations). The feature was trained in American and German English and has already been validated with German English in German ATC. For operational ATC training simulators with clean language, i.e. fully tuned and utilising all features (such as context), Cyrus has an accuracy of 96%. No figures are yet available for the syntactic analysis algorithms.	**Integrity/Process**: The central AI component of this feature is a knowledge base, which was implemented with the help of logic programming. Essentially, the knowledge base is a mathematical formalisation in the form of mathematical/logical formulas of information contained in flight manuals, regulations and the expertise of subject matter experts, as well as the current state of the flight context. For example, it contains rules such as 'An engine fire requires an immediate landing'; 'A turbojet has an actual landing distance of Y metres'; 'If the runway is wet, increase the required landing distance by a factor of X'; 'The Madrid METAR predicts rain', etc. When an event occurs, the system 'justifies' the formulae and assesses whether the event has an impact on the current flight strategy and, if so, makes recommendations on how to deal with the situation. For example, if an engine fire is observed, the system recommends landing as quickly as possible at a nearby airport, e.g. on a runway that is sufficiently long given the current outage. **Ability/performance**: The AI component of this feature is 'based' on the information that the knowledge engineer has manually coded in the knowledge base. It is therefore an approach that is orthogonal to machine learning, which conversely 'learns' from data. This has advantages and disadvantages. On the one hand, the knowledge base always provides a solution that follows logically from the coded formulae: The results are verifiable and explainability is a free by-product of reasoning. On the other hand, this technique is rather rigid: what is not present in the knowledge base cannot be recognised and explained. The knowledge base of this feature mainly contains formulas that are relevant for validation.	**Integrity/Process**: This feature uses a machine learning model. Machine learning means that algorithms learn statistical patterns and correlations from sample data instead of being explicitly programmed. As the amount of data increases and the model parameters are optimised, the prediction or classification capabilities of the model improve for new, unknown inputs. This feature was trained on a large dataset of traffic, weather, NOTAM, flight and aircraft information and predicts the STAR and runway based on these parameters. An additional layer of explainability is used to determine the parameters that carry the most weight in the choice. **Ability/performance**: The feature was trained using five years of data from Frankfurt Airport and was validated with an additional year of test data. In an earlier validation with flights from MAD to FRA, the prediction (the estimated combination of STAR and runway with the highest probability) matched the combination actually cleared by ATC in 71% of the tests.

Table A3. Flight details during tutorial flight segments and experimental flights in the blue and yellow version.

	Tutorial	Blue flight	Yellow flight
Technical Problems known before take-off (minimum equipment list 'MEL' items)	None	Failure of left engine generator restricts max. flight level to 330	Mach Number not displayed in primary flight display restricts max. flight level to 250
Time of day, weather and visibility conditions	Daytime, weather and visibility good	Daytime, weather and visibility good	Daytime, weather and visibility good
Taxi	Airport Torino-Caselle (LIMF): taxi from terminal to runway 18	Taxi from terminal complex T123 to RWY 36 R via taxiways M and B; route includes one change of taxi instructions	Taxi from terminal complex T123 to RWY 36 R via taxiways M and EC; route includes one change of taxi instructions
Cruise	Cruise flight between Athens International Airport (LGAV) and Madrid-Barajas Airport (LEMD) with severe turbulence intersecting the trajectory	Weather radar fails Thunderstorm area intersects flight trajectory	Bleed air leakage detected; results in wing anti-ice failure Icing area intersects flight trajectory
Approach	Two predictions from Copenhagen Airport (EKCH) to Frankfurt am Main Airport (EDDF) and from Athens International Airport (LGAV) to Frankfurt am Main Airport (EDDF)	Flight plan based on briefing document: EMPA5A, RWY 25 L RWY direction in use in reality (based on current weather conditions): 07	Flight plan based on briefing document: IMCO1D, RWY 07 R RWY direction in use in reality (based on current weather conditions): 25

Table A4. Group characteristics means and standard errors together with the statistics on group differences between early and late transparency subjects.

Characteristic	Early transparency	Late transparency	Group comparison
Number of pilots	11	10	
Age in years	41.64 (3.81)	38 (3.38)	$W = 66.5$, $p = 0.417$
Flight Hours	8690.91 (2715.16)	6520 (1343.67)	$W = 55.5$, $p = 0.972$
Captains / First Officers	3 / 8	3 / 7	$X^2 = 0.019$, $p = 0.890$
Risks Feature Taxi & Takeoff	4.55 (0.49)	3.89 (0.51)	$W = 61.5$, $p = 0.349$
Risks Feature En Route	12.91 (1.28)	10.22 (1.15)	$W = 70$, $p = 0.117$
Risks Feature Approach & Landing	5.18 (0.7)	2.78 (0.46)	$W = 83$, $p = 0.008$
Attitudes	28.55 (1.96)	28.44 (2.07)	$W = 47.5$, $p = 0.879$
Attainment	14.91 (1.12)	14.78 (0.92)	$W = 55$, $p = 0.673$
Utility	14.55 (0.95)	13.67 (0.96)	$W = 55$, $p = 0.673$

Table A5. Type-III ANOVA results for the cumulative link mixed models.

Predictor	df	χ^2	p
Manipulation Check: Demand for Explainability, dimensions: perceived own expertise, demand for explainability			
Transparency	1	4.43	**0.035**
Dimension	1	0.36	0.548
Feature	2	6.21	**0.045**
Flight Hours	1	0.03	0.861
Risks Feature Taxi & Takeoff	1	3.03	0.082
Risks Feature En-Route	1	0.04	0.838
Risks Feature Approach & Landing	1	17.92	**<0.001**
Transparency × Dimension	1	5.42	**0.02**
Tutorial Flight, dimensions: predictability, reliability			
Transparency	1	7.58	**0.006**
Dimension	1	25.01	**<0.001**
Feature	2	38.60	**<0.001**
Flight Hours	1	2.89	0.089
Risks Feature Taxi & Takeoff	1	2.30	0.13
Risks Feature En-Route	1	13.65	**<0.001**
Risks Feature Approach & Landing	1	4.17	**0.041**
Transparency × Dimension	1	0.28	0.594
Experimental Flight, dimensions: predictability, reliability, ease_of_use, intention_of_use, communication_style			
Transparency	1	8.34	**0.004**
Dimension	4	19.35	**<0.001**
Feature	2	15.15	**<0.001**
Flight Hours	1	10.02	**0.002**
Risks Feature Taxi & Takeoff	1	0.27	0.602
Risks Feature En-Route	1	4.18	**0.041**
Risks Feature Approach & Landing	1	2.19	0.138

(*Continued*)

Table A5. Continued.

Predictor	df	χ^2	*p*
Transparency × Dimension	4	10.60	**0.031**
Transparency × Feature	2	4.75	0.093
Dimension × Feature	8	18.29	**0.019**
Transparency × Dimension × Feature	8	8.64	0.373
Change between Tutorial and Experimental Flight, dimensions: predictability, reliability			
Tutorial Score	4	216.72	**<0.001**
Transparency	1	6.49	**0.011**
Dimension	1	5.76	**0.016**
Feature	2	26.71	**<0.001**
Flight Hours	1	13.94	**<0.001**
Risks Feature Taxi & Takeoff	1	0.00	0.98
Risks Feature En-Route	1	7.63	**0.006**
Risks Feature Approach & Landing	1	9.57	**0.002**
Transparency × Dimension	1	1.74	0.187

Table A6. Results for the cumulative link mixed models.

Predictor	Estimate [confidence intervals]	Bootstrap bias (SE)	*z*	*p*
Manipulation Check: Demand for Explainability, dimensions: perceived own expertise, demand for explainability				
Transparency Early	0.85 [0.06, 1.64]	0.03 (0.41)	2.10	**0.035**
Dimension Perceived own Expertise	0.19 [−0.43, 0.82]	−0.002 (0.38)	0.60	0.548
Feature En-Route	0.38 [−0.1, 0.86]	0.008 (0.26)	1.54	0.124
Feature Approach & Landing	0.61 [0.13, 1.09]	0.034 (0.26)	2.47	**0.014**
Flight Hours	−0.03 [−0.32, 0.27]	0.003 (0.1)	−0.17	0.861
Risks Feature Taxi & Takeoff	−0.22 [−0.47, 0.03]	−0.004 (0.11)	−1.74	0.082
Risks Feature En-Route	0.01 [−0.09, 0.11]	−0.001 (0.04)	0.20	0.838
Risks Feature Approach & Landing	−0.37 [−0.54, −0.2]	−0.019 (0.08)	−4.23	**<0.001**
Transparency Early × Dimension Perceived own Expertise	1.01 [0.16, 1.86]	0.064 (0.47)	2.33	**0.02**
Tutorial Flight, dimensions: predictability, reliability				
Transparency Early	1.39 [0.4, 2.39]	0.039 (0.27)	2.75	**0.006**
Dimension Predictability	1.2 [0.73, 1.68]	0.041 (0.25)	5.00	**<0.001**
Feature En-Route	1.06 [0.68, 1.44]	0.026 (0.21)	5.42	**<0.001**
Feature Approach & Landing	−0.01 [−0.38, 0.37]	0 (0.2)	−0.03	0.979
Flight Hours	0.34 [−0.05, 0.73]	0.009 (0.07)	1.70	0.089
Risks Feature Taxi & Takeoff	0.25 [−0.07, 0.58]	0.009 (0.08)	1.52	0.13
Risks Feature En-Route	−0.26 [−0.39, −0.12]	−0.008 (0.03)	−3.69	**<0.001**
Risks Feature Approach & Landing	−0.22 [−0.44, −0.01]	−0.007 (0.05)	−2.04	**0.041**
Transparency Early × Dimension Predictability	0.17 [−0.46, 0.8]	−0.003 (0.34)	0.53	0.594
Experimental Flight, dimensions: predictability, reliability, ease_of_use, intention_of_use, communication_style				
Transparency Early	1.59 [0.51, 2.68]	0.027 (0.3)	2.89	**0.004**
Dimension Reliability	−0.81 [−1.41, −0.22]	−0.021 (0.26)	−2.68	**0.007**
Dimension Predictability	0.13 [−0.56, 0.82]	0.004 (0.3)	0.38	0.706
Dimension Communication Style	0.48 [−0.18, 1.15]	0.01 (0.32)	1.43	0.152
Dimension Intention-to-use	1.4 [0.07, 2.72]	0.038 (1.77)	2.07	**0.038**
Feature En-Route	0.29 [−0.24, 0.82]	0.008 (0.25)	1.08	0.28
Feature Approach & Landing	−0.77 [−1.32, −0.22]	−0.004 (0.3)	−2.77	**0.006**
Flight Hours	0.64 [0.24, 1.03]	0.014 (0.05)	3.17	**0.002**
Risks Feature Taxi & Takeoff	−0.09 [−0.42, 0.24]	0.002 (0.04)	−0.52	0.602
Risks Feature En-Route	−0.14 [−0.28, −0.01]	−0.003 (0.02)	−2.04	**0.041**
Risks Feature Approach & Landing	−0.16 [−0.38, 0.05]	−0.001 (0.03)	−1.48	0.139
Transparency Early × Dimension Reliability	−0.62 [−1.45, 0.21]	−0.016 (0.38)	−1.47	0.141
Transparency Early × Dimension Predictability	−0.07 [−1.04, 0.9]	−0.011 (0.48)	−0.14	0.887
Transparency Early × Dimension Communication Style	−1.29 [−2.2, −0.38]	−0.026 (0.48)	−2.78	**0.005**
Transparency Early × Dimension Intention-to-use	−1.62 [−3.42, 0.17]	−0.016 (2.24)	−1.77	0.077
Transparency Early × Feature En-Route	0.3 [−0.46, 1.06]	−0.006 (0.41)	0.77	0.44
Transparency Early × Feature Approach & Landing	0.84 [0.08, 1.61]	0.006 (0.45)	2.16	**0.031**
Dimension Reliability × Feature En-Route	0.26 [−0.59, 1.11]	0.003 (0.37)	0.60	0.547
Dimension Predictability × Feature En-Route	0.09 [−0.88, 1.07]	−0.028 (0.44)	0.18	0.854
Dimension Communication Style × Feature En-Route	−0.7 [−1.63, 0.23]	−0.023 (0.45)	−1.48	0.14
Dimension Intention-to-use × Feature En-Route	−1.77 [−3.59, 0.04]	−0.095 (2.31)	−1.91	0.056
Dimension Reliability × Feature Approach & Landing	0.43 [−0.42, 1.28]	0.007 (0.4)	1.00	0.319
Dimension Predictability × Feature Approach & Landing	−0.05 [−1.05, 0.94]	−0.01 (0.5)	−0.10	0.92
Dimension Communication Style × Feature Approach & Landing	−0.49 [−1.43, 0.44]	−0.021 (0.5)	−1.03	0.301
Dimension Intention-to-use × Feature Approach & Landing	−3.43 [−5.33, −1.53]	−0.282 (2.95)	−3.54	**<0.001**
Transparency Early × Dimension Reliability × Feature En-Route	−0.37 [−1.55, 0.81]	0.017 (0.56)	−0.61	0.542
Transparency Early × Dimension Predictability × Feature En-Route	−0.51 [−1.89, 0.87]	0.032 (0.64)	−0.72	0.469
Transparency Early × Dimension Communication Style × Feature En-Route	0.52 [−0.76, 1.81]	0.02 (0.65)	0.80	0.426

(*Continued*)

Table A6. Continued.

Predictor	Estimate [confidence intervals]	Bootstrap bias (SE)	*z*	*p*
Transparency Early × Dimension Intention-to-use × Feature En-Route	1.17 [−1.37, 3.71]	0.207 (3.13)	0.90	0.366
Transparency Early × Dimension Reliability × Feature Approach & Landing	−1.13 [−2.32, 0.05]	−0.017 (0.63)	−1.88	0.061
Transparency Early × Dimension Predictability × Feature Approach & Landing	−0.93 [−2.34, 0.47]	−0.007 (0.81)	−1.30	0.193
Transparency Early × Dimension Communication Style × Feature Approach & Landing	0.05 [−1.25, 1.35]	0.021 (0.73)	0.08	0.938
Transparency Early × Dimension Intention-to-use × Feature Approach & Landing	1.64 [−0.94, 4.21]	0.209 (3.36)	1.25	0.213
Change between Tutorial and Experimental Flight, dimensions: predictability, reliability				
Tutorial Answer 2	−1.87 [−2.65, −1.08]	−0.11 (0.51)	−4.66	**<0.001**
Tutorial Answer 3	−4.04 [−4.88, −3.19]	−0.196 (0.51)	−9.38	**<0.001**
Tutorial Answer 4	−5.43 [−6.34, −4.53]	−0.247 (0.54)	−11.74	**<0.001**
Tutorial Answer 5 (strongly agree)	−7.24 [−8.35, −6.13]	−0.315 (0.65)	−12.79	**<0.001**
Transparency Early	1.46 [0.61, 2.31]	0.002 (0.33)	3.36	**<0.001**
Dimension Reliability	−0.59 [−1.08, −0.11]	−0.019 (0.25)	−2.40	**0.016**
Feature En-Route	0.26 [−0.13, 0.66]	−0.006 (0.21)	1.33	0.185
Feature Approach & Landing	−0.74 [−1.13, −0.35]	−0.035 (0.22)	−3.71	**<0.001**
Flight Hours	0.57 [0.27, 0.87]	0.027 (0.07)	3.73	**<0.001**
Risks Feature Taxi & Takeoff	0 [−0.26, 0.25]	0.016 (0.07)	−0.03	0.98
Risks Feature En-Route	−0.15 [−0.26, −0.04]	−0.008 (0.03)	−2.76	**0.006**
Risks Feature Approach & Landing	−0.27 [−0.43, −0.1]	−0.004 (0.06)	−3.09	**0.002**
Transparency Early × Dimension Reliability	−0.43 [−1.07, 0.21]	0.019 (0.33)	−1.32	0.187