

# Explainable artificial intelligence in air traffic control: effects of expertise on workload, acceptance, and usage intentions

Received: 2 August 2025

Accepted: 26 November 2025

Published online: 24 January 2026

Cite this article as: Cartocci G., Veyrié A., Cavagnetto N. *et al.* Explainable artificial intelligence in air traffic control: effects of expertise on workload, acceptance, and usage intentions. *Brain Inf.* (2025). <https://doi.org/10.1186/s40708-025-00287-6>

Giulia Cartocci, Alexandre Veyrié, Nicola Cavagnetto, Christophe Hurter, Augustin Degas, Ana Ferreira, Mobyen Uddin Ahmed, Shahina Begum, Shaibal Barua, Bianca Maria Serena Inguscio, Vincenzo Ronca, Gianluca Borghini, Gianluca Di Flumeri, Fabio Babiloni & Pietro Aricò

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

# Explainable Artificial Intelligence in Air Traffic Control: Effects of Expertise on Workload, Acceptance, and Usage Intentions

Giulia Cartocci<sup>1,2\*</sup>, Alexandre Veyrié<sup>3\*</sup>, Nicola Cavagnetto<sup>4</sup>,  
Christophe Hurter<sup>3</sup>, Augustin Degas<sup>3</sup>, Ana Ferreira<sup>4</sup>,  
Mobyen Uddin Ahmed<sup>5</sup>, Shahina Begum<sup>5</sup>, Shaibal Barua<sup>5</sup>,  
Bianca Maria Serena Inguscio<sup>6</sup>, Vincenzo Ronca<sup>2,6</sup>,  
Gianluca Borghini<sup>1,2</sup>, Gianluca Di Flumeri<sup>1,2</sup>, Fabio Babiloni<sup>2,7,8</sup>,  
Pietro Aricò<sup>2,6</sup>

<sup>1</sup>Department of Molecular Medicine, Sapienza University of Rome, Rome, 00185, Italy.

<sup>2</sup>BrainSigns Srl, Industrial Neurosciences Lab, Rome, 00198, Italy.

<sup>3</sup>Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, Toulouse, 31400, France.

<sup>4</sup>Deepblue Srl, Rome, 00185, Italy.

<sup>5</sup>Artificial Intelligence and Intelligent Systems Group, Mälardalen University, Västerås, 72220, Sweden.

<sup>6</sup>Department of Computer, Control, and Management Engineering, Sapienza University of Rome, Rome, 00185, Italy.

<sup>7</sup>Department of Physiology and Pharmacology, Sapienza University of Rome, Rome, 00185, Italy.

<sup>8</sup>College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, 310005, China.

\*Corresponding author(s). E-mail(s): [giulia.cartocci@uniroma1.it](mailto:giulia.cartocci@uniroma1.it);  
[alexandre.veyrie@enac.fr](mailto:alexandre.veyrie@enac.fr);

Contributing authors: [nicola.cavagnetto@dblue.it](mailto:nicola.cavagnetto@dblue.it);  
[christophe.hurter@enac.fr](mailto:christophe.hurter@enac.fr); [augustin.degas@enac.fr](mailto:augustin.degas@enac.fr);  
[ana.ferreira@dblue.it](mailto:ana.ferreira@dblue.it); [mobyen.uddin.ahmed@mdu.se](mailto:mobyen.uddin.ahmed@mdu.se);  
[shahina.begum@mdu.se](mailto:shahina.begum@mdu.se); [shaibal.barua@mdu.se](mailto:shaibal.barua@mdu.se);  
[bianca.maria.inguscio@uniroma1.it](mailto:bianca.maria.inguscio@uniroma1.it); [vincenzo.ronca@uniroma1.it](mailto:vincenzo.ronca@uniroma1.it);

[gianluca.borghini@uniroma1.it](mailto:gianluca.borghini@uniroma1.it); [gianluca.diffumeri@uniroma1.it](mailto:gianluca.diffumeri@uniroma1.it);  
[fabio.babiloni@uniroma1.it](mailto:fabio.babiloni@uniroma1.it); [pietro.arico@uniroma1.it](mailto:pietro.arico@uniroma1.it);

### Abstract

Explainability is crucial for establishing user trust in Artificial Intelligence (AI), particularly within safety-critical domains such as Air Traffic Management (ATM) and Air Traffic Control (ATC). This study empirically investigates the effects of Explainable AI (XAI), specifically HeatMap-based visual explanations, on cognitive workload, user acceptance, and intention to use AI-driven decision-support systems among Air Traffic Control Officers (ATCOs). Despite significant theoretical advancements in the broader XAI domain, empirical evidence addressing the specific impact of visual explanations on human-AI interactions in safety-critical environments like ATC remains limited. To address these critical gaps, an experimental comparison was conducted between explainable (HeatMap) and non-explainable (BlackBox) AI conditions, involving two user groups: expert and student ATCOs. Both objective neurophysiological measures (Electroencephalography) and subjective questionnaires were employed to capture comprehensive user responses. Key findings revealed that the presence of visual explanations significantly reduced cognitive workload and enhanced users' willingness to adopt the AI system, regardless of participants' level of expertise. However, explicit perceptions of AI's impact on work performance were predominantly influenced by expertise, with less experienced controllers reporting a greater perceived impact than their expert counterparts. By combining objective neurometrics with subjective user assessments, this research advances methodological rigor in evaluating human-AI interactions and highlights the importance of tailored, user-centric explanations. These findings directly contribute to practical guidelines for designing cognitively compatible and trustworthy AI tools in ATC, providing nuanced insights for targeted training and deployment strategies based on user expertise.

**Keywords:** Explainable Artificial Intelligence, Air Traffic Control, Electroencephalography, Workload, Expertise

## 1 Introduction

Artificial intelligence (AI) has become a pivotal component in modern Air Traffic Management (ATM) and Air Traffic Control (ATC), driven by the growing complexity and volume of global air traffic. Traditional systems, reliant on human controllers for real-time decision-making, are increasingly augmented by AI algorithms that support predictive analytics, conflict detection, and dynamic routing [1–4]. These systems process massive streams of radar, weather, and flight data far faster and more accurately than human operators can, enabling earlier detection of potential conflicts and more efficient use of airspace. As a result, Air Traffic Control Officers (ATCOs), whose main duty is to prevent the collision of airplanes, are shifting from direct tactical control toward supervisory roles, where they oversee and intervene in the actions of

automated systems [5–8]. This transition introduces new cognitive demands, requiring controllers to maintain situational awareness in the presence of opaque or rapidly changing AI-driven decisions [9–11]. Integration challenges remain, especially in balancing human judgment with algorithmic recommendations, but the trend toward collaborative human-AI control paradigms is reshaping the operational and cognitive landscape of ATC [12–15].

The integration of AI in critical decision-making environments, such as ATC, necessitates ensuring that AI-driven recommendations are both transparent and comprehensible to human operators. Transparency is particularly crucial in these safety-critical fields, where understanding the rationale behind AI-generated solutions can directly affect human trust, acceptance, and operational performance [16–18]. This requirement is emphasized by regulatory authorities such as the European Union Aviation Safety Agency (EASA) who establishes that: “A crucial component of achieving Trustworthy AI is transparency which encompasses three elements: traceability, explainability and open communication about the limitations of the AI system”. Consequently, AI systems must deliver more than just technical performance; they must also offer strong explainability. Furthermore, EASA defines explainability as the “capability to provide the human with understandable, reliable, and relevant information with the appropriate level of detail and with appropriate timing on how an AI application produces its results”.

Explainable Artificial Intelligence (XAI), encompassing methods that clarify AI decision-making processes, has thus emerged as an essential research domain, addressing concerns around interpretability, reliability, and trustworthiness of AI systems [19–26]. In a recent and comprehensive survey of the emerging field of XAI, aiming to address the opacity of complex AI systems, Arrieta et al., (2020) proposed a unified, user-centered definition of explainability and presented two structured taxonomies: one for general machine learning models and another specifically for deep learning approaches. They highlighted key challenges in achieving effective XAI, including the trade-off between interpretability and model performance, the absence of standard evaluation metrics, and the need for context-aware, user-adapted explanations [20]. In another recent review, Degas et al., (2022) identified two critical gaps in the current landscape of XAI for ATM. On one hand, they highlighted the urgent need for a shift toward user-centric XAI, where systems are not only technically capable but also designed to align with the needs, workflows, and mental models of ATCOs [21]. Much of the existing research remains system-focused, prioritizing what AI can do, rather than what users actually need to understand [16, 25].

Explanations are often chosen arbitrarily by researchers, ranging from global overviews (e.g., full decision trees) to local snippets (e.g., individual decision branches), with little input from end users. On the other hand, they carried out attention to a persistent confusion between the terms transparency and explainability, a conceptual ambiguity that continues to hinder progress in XAI. Although the difference may seem nuanced, clarity is essential. Transparency refers to system-level openness (e.g., exposing internal mechanics), while explainability focuses on user-facing clarity and relevance (e.g., offering contextualized, understandable insights) [27]. In this way, a

tool can be transparent yet not explainable, such as revealing a neural network's architecture without making its outputs interpretable to an ATCO, or explainable without being transparent, as in summarizing a system's decision while hiding proprietary algorithms. These concepts, though related, are not interchangeable; both are crucial to fostering trust in AI.

In this paper, we draw a clear distinction between transparency and explainability, as these terms are often blurred in related work. Transparency refers to an inherent property of a system, whereby its internal functioning can be directly inspected and understood without additional mediation (e.g., rule-based systems or decision trees). Explainability, by contrast, denotes post-hoc methods or interfaces that provide users with accessible reasons or visualizations of the system's outputs when the underlying model is not inherently transparent (e.g., saliency maps, heatmaps, or surrogate models). The present study focuses on explainability, since the AI model employed for conflict detection and resolution is not transparent by design but requires additional explanatory interfaces to make its recommendations interpretable to ATCOs.

Despite significant theoretical advances in the broader domain of XAI, empirical evidence regarding the effectiveness of specific visual explanation methods, such as HeatMap-based visualizations, remains scarce within the aviation sector. Given the highly dynamic and cognitively demanding nature of ATC tasks [28–31], understanding how such visual explanations influence cognitive workload, user acceptance, and intention to use AI-driven tools is paramount. Prior studies indicate that visual explanatory methods can serve as cognitive anchors, facilitating rapid comprehension and efficient decision-making [32, 33], although their effectiveness may depend significantly on the user's level of expertise and pre-existing operational routines [34–36]. In support of this view, [22] recently presented a framework for integrating XAI into ATC, highlighting a crucial shift in the design of XAI systems for ATCOs. Rather than assuming that explanations are universally beneficial, the authors argue that explanations should be tailored to the specific instances when and why ATCOs actually need them. Their study, grounded in operational goals of ATM, found that explanations are particularly valuable when ATCOs must document decisions or resolve conflicts between their judgment and AI-generated advisories. However, when the AI system aligns with the controller's expectations, explanations are often unnecessary and may even contribute to cognitive overload. This user-centered approach emphasizes that effective XAI in ATM must be context-aware, delivering relevant explanations only when they enhance trust, transparency, or accountability in human-AI collaboration [26]. These findings underscore the importance of moving beyond system-level design to examine how explanatory interfaces concretely affect ATCOs' cognitive experience and behavioral intentions, particularly in terms of mental workload, system acceptance, and willingness to adopt AI support tools in practice.

Mental workload, user acceptance, and intention to adopt AI-driven systems are interconnected human factors critical to the successful integration of AI in ATC. Mental workload, defined as the cognitive resources required by a human operator to perform specific tasks, directly impacts both performance and safety in high-stakes domains such as aviation [29, 37]. Mental workload is a central consideration in the

adoption of AI-based decision support systems in ATC, as the integration of automation can both alleviate and amplify cognitive demands depending on task complexity, system transparency, and user expertise [28–30]. Excessive cognitive load can lead to decreased situational awareness, delayed reaction times, and increased error rates, underscoring the necessity of designing AI support tools that effectively mitigate, rather than exacerbate, cognitive demands on ATCOs [9, 31, 38]. Consequently, AI systems must be designed not only for technical reliability but also for cognitive compatibility with human operators.

This has direct implications for system acceptance and intention to use, two key dimensions in technology adoption frameworks such as the Technology Acceptance Model (TAM) [39, 40]. Furthermore, the acceptance of AI-driven decision aids among ATCOs hinges significantly on the perceived utility, ease of use, and reliability of the system [34, 41]. Trust in AI systems, shaped by transparency and explainability, fundamentally affects user acceptance and determines whether controllers incorporate AI recommendations into their operational routines or dismiss them entirely [17, 21]. Lastly, intention to use AI tools is strongly predicted by these acceptance factors and the controller’s subjective evaluation of the system’s effectiveness in supporting their operational tasks. Research consistently highlights that ATCOs’ intention to adopt AI systems increases when explanations provided by the AI align with their cognitive expectations, operational experiences, and professional judgment [22, 33].

Hence, addressing these dimensions collectively, mental workload, acceptance, and usage intention, is vital for fostering effective human-AI teaming in the safety-critical environment of ATC. Research shows that when AI tools are perceived as supportive, particularly when their outputs are transparent and cognitively aligned with operator needs, they are more likely to be accepted and incorporated into routine ATC workflows [8, 27]. Conversely, if AI decisions are perceived as opaque or cognitively burdensome, operators may disengage, resist adoption, or default to manual control, particularly in high-stakes situations [42]. In this context, explainability serves as a mediating factor between system complexity and user acceptance by reducing cognitive uncertainty and enhancing perceived usefulness [20, 25]. However, explainability itself must be carefully designed; overly complex or redundant explanations may increase workload rather than mitigate it [35]. Thus, the relationship between AI explainability, cognitive workload, and intention to use is highly dynamic, requiring empirical validation within real-world operational settings such as ATC.

Electroencephalography (EEG) and derived neurometrics, quantitative indicators extracted from brain activity signals that reflect cognitive states such as mental workload, have emerged as reliable and objective tools for assessing mental effort across various contexts, enabling empirical validation of human-AI interactions under ecologically valid conditions [28, 43]. Nonetheless, studies on the differences in cognitive workload and user acceptance that may occur between experienced and inexperienced ATCO under different task difficulties are quite insufficient in the literature. Addressing these knowledge gaps, the present study empirically investigates how visual explanations, specifically HeatMap-based XAI, influence cognitive workload, user acceptance, and willingness to adopt AI support tools among ATCOs. A combination of EEG-based neurometric assessments and targeted subjective questionnaires

was used in the present study to capture both implicit and explicit user responses. Acknowledging the critical role of expertise, the current experimental design contrasts expert ATCOs instructors against less experienced student controllers, thereby enabling an in-depth exploration of how expertise modulates responses to visual AI explanations. In this way, this research work seeks not only to extend existing theoretical frameworks by empirically validating the impact of visual explanations on human factors in AI-supported ATC systems but also to address critical gaps identified by prior literature reviews and expert recommendations related to XAI systems in ATC [20, 25, 34].

Based on the literature, three research questions have been formulated to address key gaps identified in the previous sections:

- RQ1: To what extent does XAI (HeatMap-based visualization) influence the cognitive workload of ATCOs compared to a non-explainable (black-box) AI system?
- RQ2: How does the provision of XAI affect user acceptance and intention to use AI decision-support tools in ATC?
- RQ3: Does the expertise level of ATCOs moderate the effects of XAI on workload, acceptance, and intention to use?

From these research questions, some hypotheses can be directly investigated in the present work:

- H1: XAI (visual explanations using HeatMap) reduces cognitive workload (measured via EEG neurometrics) compared to a non-explainable (black-box condition) AI system.
- H2: XAI increases users' intention to use the AI tool compared to a non-explainable AI system (BlackBox).
- H3: Acceptance of AI tools (measured via EEG neurometrics and subjective measures) is higher among student ATCOs than expert ATCOs.
- H4: The impact of XAI on intention to use and acceptance depends significantly on the user's expertise.

## 2 Materials and Methods

### 2.1 Participants

In total, 21 participants were recruited at the École Nationale de l'Aviation Civile (ENAC, Toulouse, France), the first European training school for ATCOs and pilots, to participate in the study. Participants were divided into two groups based on expertise: "Expert" and "Student" air traffic controllers (Table 1).

The Expert group included 11 participants (3 female; mean age: 41 years, range 34–51). This group was mostly composed of ENAC ATCO instructors (seven), with additional participants being former ATCOs now working in research (two) or in training/formation roles (two). On average, the experts had 15 years of operational experience, primarily in en-route control, though several also had approach and tower backgrounds.

The Student group included 10 participants (4 female; mean age: 22 years, range 20–26). These were ENAC trainees enrolled in the final stage of their formation, just before being assigned to an operational control centre. This ensured that, while not yet licensed, they had completed nearly all academic and simulator training and were able to perform conflict-resolution tasks in a realistic experimental setting.

All participants were healthy volunteers with no history of neurological or psychiatric disorders, and with normal or corrected-to-normal vision. Exclusion criteria included refusal or inability to provide informed consent or the presence of medical conditions incompatible with EEG recording. The study was approved by the Research Ethics Committee of Toulouse University and conducted in accordance with the Helsinki Declaration and institutional requirements. All participants provided written informed consent prior to participation and were free to withdraw at any time.

**Table 1** Summary of experimental participants involved in the study.

Participants	Description
Expert group (Professional ATCOs)	<ul style="list-style-type: none"> <li>• Total of 11 professional ATCOs</li> <li>• 3 female (27%) and 8 male (73%)</li> <li>• Mean age of 41 years (ranging between 34–51 years old)</li> <li>• Mean 15 years of working experience</li> <li>• Operational background: Only 2 participants did not have experience in En route ACC, but both had Approach ACC and Tower ATC experience</li> </ul>
Student group	<ul style="list-style-type: none"> <li>• Total of 10 student ATCOs</li> <li>• 4 female (40%) and 6 male (60%)</li> <li>• Mean age of 22 years (ranging between 20–26 years old)</li> </ul>

## 2.2 Experimental Task

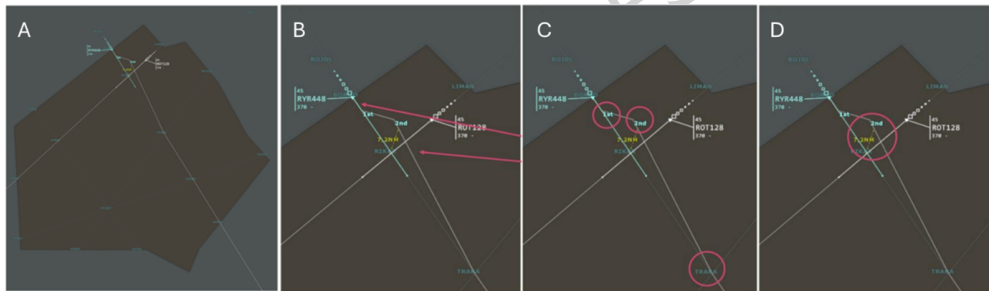
One of the principal functions of ATC is to ensure that aircraft do not conflict, that is, do not approach each other sufficiently close to give rise to the possibility of collision. In this study, a Conflict Detection & Resolution Task [44, 45] was performed using a Genetic Algorithm [46]. Genetic Algorithms (GA) are metaheuristic-based algorithms inspired by the process of natural selection that belong to the larger class of evolutionary algorithms and are commonly used to generate high-quality solutions to optimization and search problems by relying on biologically inspired operators such as mutation, crossover and selection [47–50]. As a population and evolutionary-based metaheuristic, the Genetic Algorithm (GA) means that a GA tries to iteratively improve candidate solutions according to some predefined criteria. In the case of conflict resolution, a candidate solution for the GA is a set of modified or not trajectories.

Candidate solutions forming the population are evaluated in function of three criteria: the duration of the conflicts, if any; the length of the trajectories; and the number of changes of direction (i.e. order that must be given to implement this candidate solution). Once the GA has evaluated all candidate solutions in the population, it selects a set of candidate solutions, mostly the best, but also other candidate solutions to better explore the solution space. The algorithm then applies a set of mutation and crossover operations in an attempt to enhance the population and to possibly converge toward one of the optimal solutions [46, 49].

To sum up, in the Conflict Detection and Resolution task, the GA explores the possible solutions for a given conflicting situation between aircraft and extracts one solution which can be qualified as the “best” one with the given optimization criteria (e.g., number of actions, length of the trajectory, and number of orders). In order to provide explanations for the proposed solution, two different types of data presentation were used as follows:

**BlackBox (BB):** Simple visualization which only displays the proposed solution by the GA, enhanced by instructions to proceed (Figure 1). Aeroplane trajectories are coloured differently. The minimal distance between aeroplanes is computed and displayed in yellow. The control orders that must be given by the ATCO to the different aeroplanes are placed along the trajectory, as well as their ordering (First, Second,... ).

This data presentation is not an explanation by itself, but the simple data presentation of the “best” solution the GA managed to extract. Compared to the existing system, the BlackBox data representation directly provides a solution to a detected conflict, while the system currently used only displays the detected conflicting aircraft without further information to solve it.



**Fig. 1** In the BlackBox (BB) explanation, the subject is presented with the trajectory of the aircraft and any trajectory changes (A). When the trajectory of an aircraft is modified, the initial trajectory of the aircraft appears in colour (in this case, cyan) and the modified trajectory appears in white (B). Each aircraft and its course changes have a unique associated colour. In this example, the trajectory of flight RYR448 is modified, while the trajectory of the ROT128 aircraft remains unchanged. The RYR448 aeroplane must first modify its trajectory at point 1 (First), then at point 2 (Second). It continues its trajectory after reaching the TRAKA point (C). Finally, the minimum distance between the two aeroplanes will be shown in yellow (in this case, 7.2 Nautical Meters). When the distance is less than 7 NM, it is indicated in orange, and the associated trajectories are displayed in yellow-green (D).

Heatmap (HM): To make the rationale behind the proposed solution more explicit, the presentation highlighted both the “best” solution identified by the GA and the broader solution space explored, including both favorable and unfavorable trajectories. Heatmaps of the explored trajectories were displayed to show how aircraft paths could be safely modified (Figure 2).

For example, the operator could see that AFR3218 could either maintain its current trajectory or turn left (although this would be less efficient), while KLM1258 and EZY208 could not continue on their original trajectories and were required to turn left as the only safe option. In addition, the timing of possible trajectory changes was conveyed by the extent of the “safe zone” (green area) before it transitioned into a “dangerous zone” (red area). This visualization was generated from a cumulative density map of explored candidate solutions, where each trajectory was convolved with a Gaussian kernel and then accumulated. Such a technique helps visually define areas, also known as contour maps [51], and supports reasoning about both “good” and “bad” solutions.

The Heatmap was specifically designed to create an uncluttered view of the possible trajectory modifications, filtering out less relevant candidate solutions while emphasizing those that help controllers understand why a given solution was proposed and why not others. For instance, it enables contrastive questions such as: “Can IBE752 turn left instead of continuing straight, despite the AI suggesting a right turn?” (it should not), or “What happens if I extend the modification of the IBERIA trajectory for a longer period of time?” (it would not provoke any conflict with FIN2655).

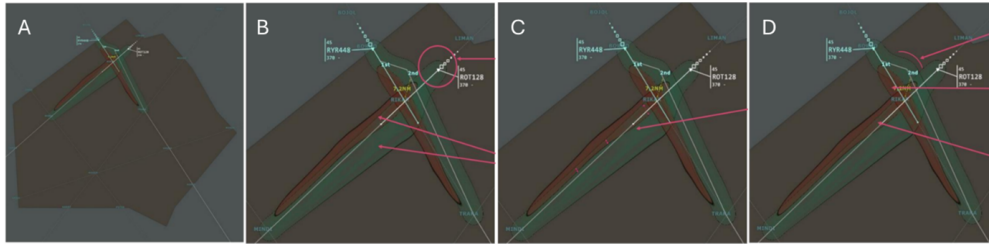
Heatmaps are also widely used in Explainable AI (XAI), for example in techniques such as LIME, LRP, and SHAP, to illustrate the contribution of input features to model outputs [52–54]. In the ATC domain, this approach was chosen because it provides controllers with an intuitive visual anchor that highlights both safe and unsafe envelopes while aligning with the strong visual culture of ATC operations. Unlike textual explanations, which risk monopolizing attention and disrupting situational awareness in fast-paced environments, visual explanations such as heatmaps can be integrated directly into radar-like displays, offering cognitively efficient and operationally relevant support.

### 2.3 Experimental Setup

Scalp EEG was recorded with the wearable Mindtooth EEG headset (Figure 3, Brain Products GmbH, Germany<sup>1</sup>) consisting of 8 electrodes positioned by following the International 10-20 EEG System (AFz, AF3, AF4, AF7, AF8, Pz, P3, P4), referenced to the right mastoid and grounded to the left one. The wet electrodes of the Mindtooth headset consist of open-cell, hydrophilic and highly absorbent cylindrical sponges. They harden when dry and become soft and expand when wet. The porous material is designed to absorb aqueous electrolyte solutions (1-2 % sodium chloride solutions are used as electrolytes). The electrolyte solution facilitates the establishment of an easy electrical connection to the participant’s skin. The impedance value was reduced to 100 k $\Omega$  before the start of the recording [55, 56].

---

<sup>1</sup>[www.mindtooth-eeg.com](http://www.mindtooth-eeg.com)



**Fig. 2** In the HeatMap (HM) explanation, the algorithm tests different trajectories and detects any associated conflicts (A). The HeatMap shows to the subject for each aircraft both information: i) an envelope of “good modifications” of trajectory (in green,  $> 7\text{NM}$ ) and ii) an envelope of “bad modifications” of trajectory (in red,  $< 7\text{NM}$ ). Thus, the HeatMap the subject sees is a representation of the accumulation of the number of trajectories explored in the same place (B). To identify the plane corresponding to the green envelope, the subject can refer to the starting point of the zone and the current location of the plane. The red envelope follows the path of the associated green envelope (C). Finally, when two zones (green/red) overlap, this means that the plane can pass through this zone (in this case, ROT128) provided that the second plane (RVR448) changes its trajectory sufficiently (D).



**Fig. 3** Mindtooth EEG headset, employed during the experiments.

The participants were placed in the ACHIL En-Route control setting, ATC simulators facilities in ENAC, Toulouse, France. The control screen was either displaying the simulation, the solution and level of explanation, the drawing, or the survey (Figure 4).

## 2.4 Experimental Design

The experimental design (Figure 5) aimed at testing primarily the different levels of explainability (BlackBox – No Explainability, and HeatMap – Explainability), while decreasing as much as possible any risk of bias, maximizing the quality of neuro-physiological measures, and keeping the experiment short enough to not induce other



**Fig. 4** Validation setup for the neurophysiological measures.

confounding factors, such as mental fatigue. Before starting the experiment, each participant underwent an initial briefing and training session on the platform and the neurophysiological setting.

Every level of explanation was then tested twice. The first scenario was used as a warm-up (Warmup Scenario), while the following one (Experimental Scenario) was used to gather data for the analyses included in this paper, by randomizing the presentation of explainability. In particular, the researchers used a Latin square to randomize the order in which the level of explanation was presented to the participants to avoid any bias linked to the order of presentation, fatigue, or scenario.

At the end of each scenario, participants completed questionnaires and were briefly interviewed about their experience (see Questionnaire and debriefing section below for additional details). At the end of the experiment, the researchers conducted a qualitative interview with the participants to gather more insights into the presented XAI support. The experimental plan was designed to keep the total session under 1.5 hours per participant.

## 2.5 Collected Data

### 2.5.1 Questionnaire Measures

At the end of each experimental scenario, participants were asked to complete a self-report, ad-hoc 3-question questionnaire assessing their agreement with the proposed solution and its comprehensibility, using a 5-point Likert Scale. At the end of each condition, the participants were also asked to fill in a self-report ad-hoc questionnaire to assess the willingness to use the XAI solution, the trust towards the support, the situational awareness gained through the usage of the support, the acceptability of

Briefing, training on platform, neurophysiological settings		30 min
C1: Black box condition (or Heatmap)	Warmup Scenario	20 min
	Experimental Scenario	
	Questionnaire and debriefing	
Break		10 min
Session Time: 60 min		
C2: Heatmap condition (or Black box)	Warmup Scenario	20 min
	Experimental Scenario	
	Questionnaire and debriefing	
Final questionnaire and interview		10 min
Session Time: 90 min		

**Fig. 5** Experimental protocol and related timing.

the solution, and the general perceived impact on the work performance, on a 5-step Likert Scale (see Appendix A). Correlation analyses were conducted between related questionnaire items to verify whether they measured the same construct. Strong correlations justified aggregating items into composite indices (e.g., willingness-to-use and work performance), thereby improving reliability and simplifying statistical comparisons (Figures 6 and 7).

Correlation Matrix

		I would like to use this tool in the future	I like the new decision support interface
I would like to use this tool in the future	Pearson's r	—	
	df	—	
	p-value	—	
I like the new decision support interface	Pearson's r	0.761	—
	df	40	—
	p-value	< .001	—

Note.  $H_0$  is positive correlation

**Fig. 6** Correlation matrix for the willingness to use items (1).

Correlation Matrix

		Working with this tool would allow me to solve conflicts faster	Working with this tool would increase my accuracy in solving co	Working with this tool would improve my performance	Using this tool would make my work easier
Working with this tool would allow me to solve conflicts faster	Pearson's r	—			
	df	—			
	p-value	—			
Working with this tool would increase my accuracy in solving co	Pearson's r	0.537	—		
	df	40	—		
	p-value	< .001	—		
Working with this tool would improve my performance	Pearson's r	0.671	0.758	—	
	df	40	40	—	
	p-value	< .001	< .001	—	
Using this tool would make my work easier	Pearson's r	0.672	0.581	0.566	—
	df	40	40	40	—
	p-value	< .001	< .001	< .001	—

Fig. 7 Correlation matrix for the willingness to use items (2).

### 2.5.2 Neurophysiological Measures

Neurophysiological measures related to Mental workload and Acceptance have been obtained by a dedicated processing of the EEG signals. In the following, we provide a description of the mental states under consideration, with particular attention to the neurophysiological features associated with each of them and their related processing.

#### *Mental Workload*

In recent decades, various definitions of mental workload have been proposed, indicating that it is a complex construct due to its multifaceted nature and the interplay of multiple cognitive aspects. The measurement of mental workload is a quantitative assessment of the mental activity resulting from performing a task. Several empirical investigations have indicated that performance declines at the extremes of the workload demand continuum - that is, when the event rate is excessively high or extremely low. For these reasons, the mental workload is an important and central construct in ergonomics and human factors research. It has already been shown in various contexts that the assessment of mental workload by EEG represents a reliable and objective measurement [43, 57].

This evidence showed that the brain's electrical activities, fundamental for evaluating mental workload, are the theta and alpha EEG rhythms, located in the Pre-Frontal

Cortex (PFC) and the Posterior Parietal Cortex (PPC) regions. The theta rhythm, especially over the PFC, presents a positive correlation, i.e., it increases when mental workload increases [58]. These results are also supported by the autonomic signals of heart rate (HR) and eye blink rate (EBR), by the performances' trends and by the questionnaires for the evaluation of the perceived workload level, while the alpha rhythm, especially over the PPC, presents an inverse correlation, i.e. decreases [59]. In recent studies, it has also been demonstrated that machine learning techniques and specific brain features can be used to compute an EEG-based Workload Index, which is able to significantly discriminate workload demands during realistic tasks [60, 61].

### ***Acceptance (Approach-Withdrawal)***

Approach and withdrawal refer to the speed and ease with which a person adapts to new situations or changes. Some people adapt easily to new situations and tend to throw themselves into them, meet new people or try new things. Others, whose style is withdrawing, tend to need more time to warm up to new situations; they may hang back before they join in. That's why this concept would be very useful to assess the acceptability and consequent motivation towards a certain stimulus. A large body of research on the relation between emotion and motivation has postulated the existence of two overarching motivational systems that organize behaviour. One system involves behaviour prompted by a possible desirable outcome, whereas the other involves behaviour prompted by a possible aversive outcome. A number of such models have been proposed, including Dickinson and Dearing's, 1979 Aversive/Attractive systems, Gray's Behavioural Activation/Behavioural Inhibition systems, and Lang et al.'s (1990) [62].

[62] proposed a similar model linked to research on frontal EEG asymmetry during emotional states. He posited that frontal asymmetry was not related to the valence of an emotional stimulus but rather to the motivational system that is engaged by that stimulus. He proposed that the left PFC is involved in a system facilitating approach behaviour to appetitive stimuli, whereas the right PFC is involved in a system facilitating withdrawal behaviour from aversive stimuli. It is so possible to derive an Approach-Withdrawal neurophysiological index by the processing of the EEG signal on the PFC. Such index has been already employed in a variety of fields and conditions supporting its consistency, like experimental conditions characterized by a low-, a neutral and a high-approach positive mindset [63], the study of personality traits in children, [64], but also public service announcement [65] and neuroaesthetics [66]. This type of index has a clear impact from an ergonomic perspective in operational fields, e.g., during the testing of new HMIs and technologies, to objectively assess the degree of acceptability of the operator in front of the new solution. In the following section it is reported the processing steps necessary to calculate the different neurophysiological measures from the EEG signal.

### **2.5.3 EEG Data Processing**

The EEG signal was first band-pass filtered using a 5th-order Butterworth filter in the frequency range of 2-30 Hz. The blink artifacts have been identified and corrected by using the O-Clean algorithm [67]. EEG signals were then segmented into epochs of 1

second and if the EEG signal amplitude exceeds  $\pm 80$  ( $\mu V$ ) it was marked as artifact (threshold criterion, EEGlab [68]). From the artefact-free EEG, the Global Field Power (GFP) was calculated for each EEG frequency band of interest, on the Theta and Alpha bands. In this case, each band was defined accordingly to the Individual Alpha Frequency (IAF) value estimated for each participant [69]. Since the alpha peak is mainly prominent during rest conditions, the participants were asked to keep their eyes closed for a minute before starting the experiment. Such a condition was used to estimate the IAF value specifically for each participant. Consequently, the EEG bands are defined as:

$$\Theta = (IAF - 6) : (IAF - 2)Hz \quad (1)$$

$$\alpha = (IAF - 2) : (IAF + 2)Hz \quad (2)$$

For each mental state (Workload and Approach-Withdrawal), a subset of channels and bands of interest was chosen according to the literature [28, 66, 70–72].

The Workload neurometric was defined as:

$$WL = \frac{\Theta (\text{Frontal Channels})}{\alpha (\text{Parietal Channels})} \quad (3)$$

The Acceptance neurometric was defined as:

$$AW = \alpha (\text{Frontal } Dx) - \alpha (\text{Frontal } Sx) \quad (4)$$

### 3 Results

Regarding subjective questionnaires, items with significant positive correlations were merged to simplify analyses and interpretation. These included combining the two "acceptability" items ("I would like to use this tool in the future" and "I like the new decision support interface") into a "willingness to use" index ( $r = 0.761$ ,  $p < 0.001$ ), and merging the four "work performance" items ("Working with this tool would allow me to solve conflicts faster", "Working with this tool would increase my accuracy in solving conflicts", "Working with this tool would improve my performance", "Using this tool would make my work easier") into a composite index ( $r = 0.537$ ;  $r = 0.671$ ;  $r = 0.672$ ;  $r = 0.758$ ;  $r = 0.581$ ;  $r = 0.566$ ;  $p < 0.001$  for all the correlations) (Figure 7).

Statistical comparisons, both for qualitative and quantitative results, have been performed using a repeated measures ANOVA (CI = 0.95), showing any within (Explainability) and between (Expertise) effects. In case of significant effects, the Duncan post-hoc test has been used to perform paired comparisons.

#### 3.1 Findings from Semi-Structured Interviews

Qualitative data analysis has been conducted through semi-structured interviews in terms of Acceptance, Human Performance, Trust, Task and System Performance, XAI Applications in ATM, and Personalization of ML Algorithms. A short summary of the results is discussed below, and the details of these results can be found in [23], [27].

**Acceptance:** Students generally preferred the HeatMap (visual) condition for its clarity and appeal, though they also expressed interest in the BlackBox tool. Experts, however, were hesitant to accept AI-generated solutions due to concerns about time loss, lack of control, and being “out of the loop.” Experienced ATCOs noted that their conflict-resolution strategies—such as balancing penalties across multiple flights—were not fully captured by the algorithms. Experts ultimately favored the BlackBox solution, while students were more open to both HeatMap and BlackBox, highlighting generational and experience-based differences in acceptance and trust.

**Human Performance:** Participants valued the BlackBox solution for its simplicity, minimal disruption, and familiarity with existing en-route tools. It provided single, clear solutions and useful trajectory details. Students appreciated the visual strengths of the HeatMap solution, such as its ability to highlight potential conflicts and incorporate factors like weather. However, experts cautioned that HeatMap could clutter displays and obscure critical information. Across groups, on-demand access to the tools was seen as most useful in high-workload or complex scenarios.

**Trust:** Trust in the AI tools was limited, largely due to insufficient training and a lack of transparency. Experts stressed the need for greater explainability to build confidence, while students were more willing to accept AI solutions on par with their own. Trust was viewed as essential before operational deployment, with training and debriefings identified as key avenues to foster it.

**Task and System Performance:** The BlackBox solution was noted to facilitate faster conflict resolution, while HeatMap raised safety concerns due to clutter, cognitive tunneling, and excessive information. ATCOs worried about over-reliance on AI, potential skill degradation, and loss of vigilance over time. Mismatched AI proposals sometimes led controllers—especially students—to doubt their own solutions, risking delays or reduced situational awareness. AI support was deemed less useful for simple two-aircraft conflicts but potentially helpful in multi-aircraft scenarios.

**XAI Applications in ATM:** Higher explainability (HeatMap) was considered more valuable for training and planning tasks, where time is less critical, rather than for real-time operations. Explainability could help trainees understand algorithmic reasoning and build trust before transitioning to less visual, more operationally efficient tools. However, overreliance on AI in early training stages risks shaping novice strategies prematurely.

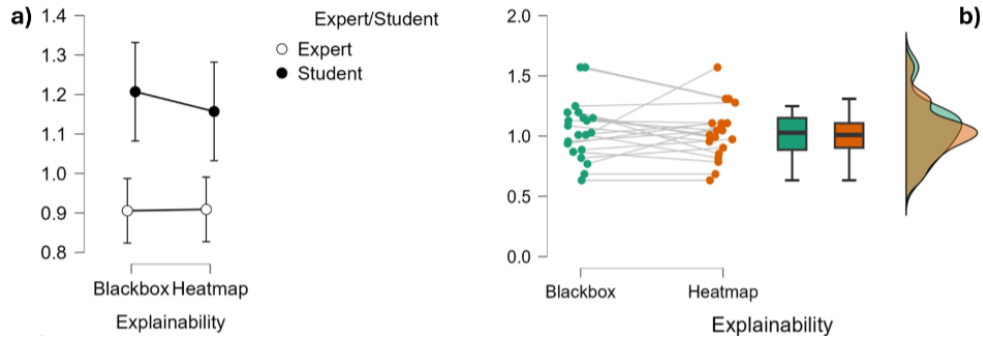
**Personalization of ML Algorithms:** Some ATCOs expressed interest in adaptive AI tools that learn individual control styles and strategies, aligning system recommendations with their decision-making preferences. Prior research confirms both common principles and individual variations in ATCO strategies, suggesting personalization could be a valuable future direction for AI support in conflict resolution.

## 3.2 Questionnaire Measures

### 3.2.1 Impact on Work Performance

On the within-subjects effects level, the main effect of Explainability was not statistically significant regarding the impact on work performances ( $F(1, 19) = 0.255$ ,  $p =$

0.620,  $\eta^2 = 0.003$ ), indicating that the presence or absence of explainability (i.e. HeatMap and BlackBox) did not significantly influence participants' responses. Similarly, the interaction between Explainability and Expertise Level was not significant ( $F(1,19) = 0.337$ ,  $p = 0.568$ ,  $\eta^2 = 0.003$ ), suggesting that the influence of Explainability did not differ between experts and students (Figure 8 and Table 2).



**Fig. 8** Top: Graphical representations of the impact of Explainability on work performances. Figure 8a shows error plots regarding the level of expertise and Explainability, while Figure 8b shows the individual points, the data distribution and the associated bar plots for explainability conditions.

Regarding the between-subjects effects level, there was a significant main effect of Expertise Level on the impact on work performance ( $F(1,19) = 16.030$ ,  $p < .001$ ,  $\eta^2 = 0.366$ ). This large effect indicates that experts and students differed significantly in their overall responses, according to their feelings about the impact on work performance, independent of the Explainability condition (Figure 8 and Table 2).

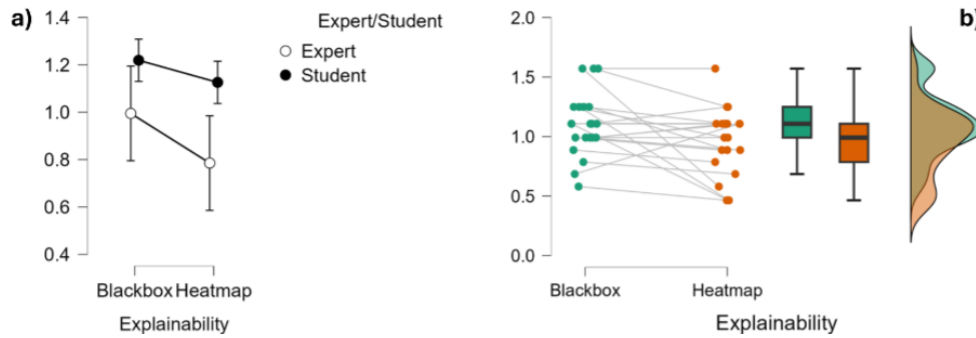
**Table 2** Statistical analyses tables displaying the impact of Explainability on work performances for students and experts, as investigated through a repeated measures ANOVA (CI=0.95), both considering within-subjects effects (top) and between-subjects effects (bottom).

Source	Sum of Squares	df	Mean Square	F	p	$\eta^2$
<i>Within-Subjects Effects</i>						
Explainability	0.006	1	0.006	0.255	.620	.003
Explainability $\times$ Expertise	0.007	1	0.007	0.337	.568	.003
Residuals	0.422	19	0.022			
<i>Between-Subjects Effects</i>						
Expertise	0.791	1	0.791	16.030	< .001	.366
Residuals	0.937	19	0.049			

*Note.* Type III Sum of Squares reported.  $\eta^2 =$  partial eta squared.

### 3.2.2 Willingness to Use

On the within-subjects effects level, there was a statistically significant main effect of Explainability on the willingness to use ( $F(1, 19) = 4.453, p = .048, \eta^2 = 0.072$ ), indicating that the presence of explainability (i.e., the HeatMap condition) significantly influenced participants' responses across conditions (Figure 9 and Table 3). However, the interaction between Explainability and Expertise Level was not statistically significant ( $F(1, 19) = 0.657, p = 0.427, \eta^2 = 0.011$ ), suggesting that the effect of Explainability did not differ significantly between experts and students.



**Fig. 9** Top: Graphical representations of the influence of Explainability on willingness to use. Figure 9a shows error plots regarding the level of expertise and Explainability while Figure 9b shows the individual points, the data distribution and the associated bar plots for Explainability conditions.

On the between-subjects effects level, there was a significant main effect of Expertise Level on the willingness to use ( $F(1, 19) = 13.210, p = 0.002, \eta^2 = 0.251$ ). This result indicates that experts and students differed significantly in their overall responses, regardless of the Explainability condition, with a large effect size (Figure 9 and Table 3).

**Table 3** Statistical analyses tables displaying the influence of Explainability on willingness to use for students and experts, as investigated through a repeated measures ANOVA (CI=0.95), both considering within subjects effects (top) and between subjects effects (bottom).

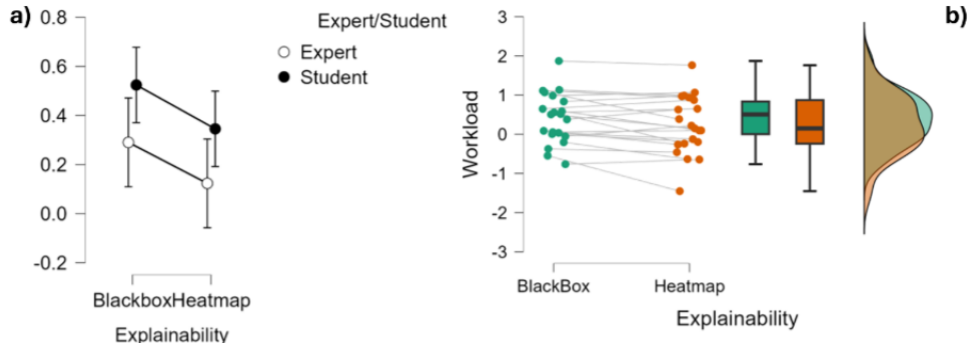
Source	Sum of Squares	df	Mean Square	F	p	$\eta^2$
<i>Within-Subjects Effects</i>						
Explainability	0.240	1	0.240	4.453	.048	.072
Explainability $\times$ Expertise	0.035	1	0.035	0.657	.427	.011
Residuals	1.024	19	0.054			
<i>Between-Subjects Effects</i>						
Expertise	0.836	1	0.836	13.210	.002	.251
Residuals	1.203	19	0.063			

Note. Type III Sum of Squares reported.  $\eta^2$  = partial eta squared.

### 3.3 Neurophysiological Measures

#### 3.3.1 Mental Workload

At the within-subjects effects level, the analysis revealed a statistically significant main effect of Explainability ( $F(1, 19) = 5.221$ ,  $p = 0.034$ ), indicating that the presence of Explainability significantly affected participants' mental workload (Figure 10 and Table 4).



**Fig. 10** Top: Graphical representations of the influence of Explainability on the measured mental workload. Figure 10a shows error plots regarding the level of expertise and Explainability, while Figure 10b shows the individual points, the data distribution and the associated bar plots for Explainability conditions.

However, the interaction between Explainability and Expertise was not significant ( $F(1, 19) = 0.006$ ,  $p = 0.939$ ), suggesting that the effect of Explainability was similar across both expert and student groups.

**Table 4** Bottom: Statistical analyses tables displaying the influence of Explainability on the measured mental workload for students and experts, as investigated through a repeated measures ANOVA (CI=0.95), both considering the within-subjects effects (top) and between-subjects effects (bottom).

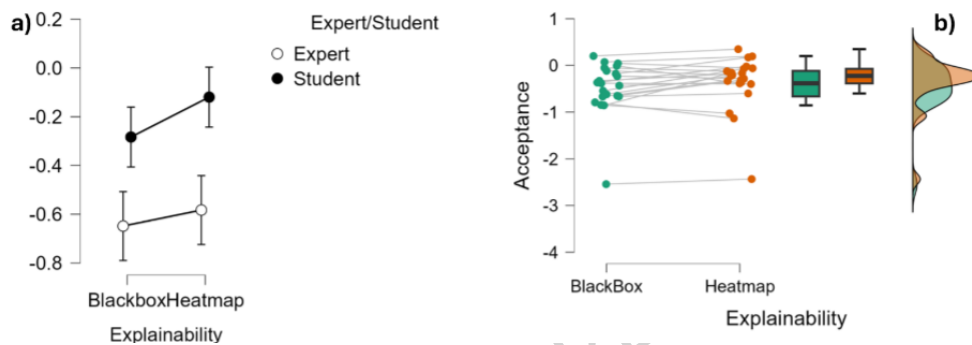
Source	Sum of Squares	df	Mean Square	F	p	$\eta^2$
<i>Within-Subjects Effects</i>						
Explainability	0.313	1	0.313	5.221	.034	.016
Explainability $\times$ Expertise	$3.580 \times 10^{-4}$	1	$3.580 \times 10^{-4}$	0.006	.939	$1.822 \times 10^{-5}$
Residuals	1.139	19	0.060			
<i>Between-Subjects Effects</i>						
Expertise	0.544	1	0.544	0.586	.454	.028
Residuals	17.652	19	0.929			

*Note.* Type III Sum of Squares reported.  $\eta^2$  = partial eta squared. Sphericity corrections not applicable for factors with two levels.

Regarding the between-subjects effects level, no significant main effect of Expertise Level on measured mental workload was found ( $F(1, 19) = 0.586, p = 0.454$ ), indicating that overall differences between experts and students were not statistically significant, regardless of the Explainability condition (Figure 10 and Table 4).

### 3.3.2 Acceptance (Approach-Withdrawal)

On the within-subjects effects level, the main effect of Explainability on the measured acceptance approached statistical significance ( $F(1, 19) = 3.696, p = 0.070, \eta^2 = 0.163$ ), suggesting a potential trend toward Explainability affecting participants' acceptance while using the system. The interaction between Explainability and Expertise was not significant ( $F(1, 19) = 0.675, p = 0.421, \eta^2 = 0.034$ ), indicating that the effect of Explainability was similar for both experts and students (Figure 11 and Table 5).



**Fig. 11** Top: Graphical representations of the influence of explainability on the measured acceptance neurometric. Figure 11a shows error plots related to the levels of Expertise and Explainability while Figure 11b shows the individual points, data distribution and associated bar plots for Explainability conditions.

Concerning the between-subjects effects level, the main effect of Expertise Level was also found not statistically significant ( $F(1, 19) = 3.116, p = 0.094, \eta^2 = 0.141$ ), though it showed a trend toward significance with a moderate effect size. In this regard, a Duncan post hoc test revealed that students exhibited an increase in acceptance (trend,  $p = 0.067$  over the HeatMap condition) (Figure 11 and Table 5).

## 4 Discussion

The present study set out to investigate how visual explanations, specifically HeatMap-based XAI, influence cognitive workload, user acceptance, and intention to adopt AI decision-support tools in the context of ATC. Recognizing the critical need for explainability in safety-critical domains, the experiment compared a black-box AI condition with a HeatMap-explanation condition across two distinct user groups: expert ATCO instructors and student controllers. By integrating objective neurometrics derived

**Table 5** Statistical analysis tables displaying the influence of explainability on the measured acceptance neurometric for students and experts, as investigated through a repeated measures ANOVA (CI = 0.95), both considering within-subjects effects (top) and between-subjects effects (bottom).

Source	Sum of Squares	df	Mean Square	F	p	$\eta^2$
<i>Within-Subjects Effects</i>						
Explainability	0.138	1	0.138	3.696	.070	.163
Explainability $\times$ Expertise	0.025	1	0.025	0.675	.421	.034
Residuals	0.708	19	0.037			
<i>Between-Subjects Effects</i>						
Expertise	1.798	1	1.798	3.116	.094	.141
Residuals	10.967	19	0.577			

Note. Type III Sum of Squares reported.  $\eta^2$  = partial eta squared.

from EEG with subjective questionnaire data, this research provided a comprehensive assessment of both implicit and explicit responses to AI explanations. The study aimed not only to evaluate the direct effects of explainability on human factors but also to explore how the level of user expertise moderates these effects. In doing so, it addressed key gaps identified in the literature regarding the operational impact of XAI in ATM and contributed empirical evidence to support user-centered design principles in the development of trustworthy AI systems [15, 73, 74].

#### 4.1 Summary of Key Findings

This study examined the influence of XAI, specifically HeatMap-based visualizations, on cognitive workload, user acceptance, and willingness to adopt AI decision-support tools within ATC. Recognizing the critical need for transparency and comprehensibility in safety-critical systems, the study assessed both explicit (subjective questionnaires) and implicit (EEG-based neurometrics) user responses, comparing experienced ATCOs with student controllers.

Subjective questionnaire data revealed no significant influence of explainability on participants' perceived impact on work performance, indicating that their explicit evaluations were unaffected by the presence or absence of visual explanations. However, a significant main effect of expertise emerged, revealing that students consistently reported a higher perceived impact of AI support on their performance than expert ATCOs. This suggests that novices, who are potentially less influenced by established routines or biases, may perceive AI assistance as more influential in their work, highlighting the important role of user experience and generational factors in shaping perceptions toward AI integration.

Furthermore, participants expressed significantly greater willingness to use the AI system when provided with explainability (HeatMap condition), compared to the black-box condition. This finding highlights the crucial role of visual explanations in promoting user acceptance of AI-based support tools. Additionally, the significant difference between expert and student ATCOs in their overall willingness to adopt the

AI tools suggests that younger, less experienced controllers may be more receptive to integrating novel technologies into their workflows, further emphasizing expertise as a key moderator in technology acceptance.

Neurometric measures, derived from EEG recordings, demonstrated a significant reduction in cognitive workload under the HeatMap-based explainability condition compared to the black-box condition, independent of user expertise. This critical finding provides objective evidence supporting the hypothesis that visual explanations effectively reduce cognitive demands by clarifying AI-generated recommendations, enabling users to allocate their cognitive resources more efficiently.

Finally, implicit measures of acceptance (Approach-Withdrawal metrics) revealed a trend towards higher acceptance levels in the presence of explainability, although this effect did not reach conventional levels of statistical significance. No significant differences were observed between the expert and student groups in this measure; however, post hoc analyses revealed a marginal trend for increased implicit acceptance among students in the HeatMap condition. This suggests that visual explanations may positively influence user perceptions at an implicit level, particularly among less experienced users, warranting further exploration.

Taken together, these results emphasize that visual explanations, such as HeatMap, positively impact user acceptance and cognitive efficiency when interacting with AI tools in ATC contexts. However, explicit perceptions of AI's impact on work performance appear more closely related to individual expertise and generational differences than to the presence of explainability alone. These findings provide a nuanced view of how explainability interacts with user characteristics and cognitive factors, preparing the ground for deeper discussion on designing AI systems that effectively align technical performance with human-centered requirements in safety-critical domains.

## 4.2 Theoretical and Practical Implications

This study found that expertise significantly influences how users perceive the impact of AI on work performance, and less experienced operators (students) reported a greater perceived impact from AI systems compared to experienced ATCOs. These results are consistent with a previous study in which the domain-specific expertise significantly impacted how users trust and engage with explainable AI systems. More experienced users tend to have lower initial trust in AI and demand clearer explanations, whereas less experienced users are more influenced by the perceived intelligence of the system itself [75]. These findings suggest that AI integration strategies in ATM should consider the users' expertise and age [36, 76]. Training programs for AI systems could be tailored to address the specific needs and biases of different user groups, ensuring that both novice and experienced operators can effectively adopt and utilize AI technologies [77]. Additionally, this underscores the importance of designing AI systems that can accommodate diverse user profiles, potentially leading to more universal adoption and improved operational performance across varying levels of expertise.

Additionally, both user experience and the type of AI explanation affect willingness to use AI systems, with less experienced operators showing a higher willingness to use AI, likely due to fewer biases towards traditional methods. The preference for BlackBox conditions among some experienced ATCOs highlights the complexity of

trust in AI systems and reinforces the previous literature [78]. Practically, this indicates the need for a flexible approach in AI design, offering both detailed explanations and more opaque options depending on the user's preference and trust level. Theoretically, it suggests a nuanced relationship between transparency and trust, where more information isn't always better. Providing excessive or complex AI explanations can lead to cognitive overload and reduce effective trust. Adjusting the amount and type of explanation based on task complexity improves both trust and performance [79]. Consequently, developing adaptive AI systems that can adjust the level of explanation based on user feedback and context could thus enhance acceptance, usability and trust of the user [73, 80].

The study also revealed that explainability significantly affects workload, independent of user expertise. Both students and experts ATCOs experienced reduced workload with the HeatMap condition compared to the BlackBox condition. This indicates that clear, visual AI explanations can alleviate cognitive load by allowing users to focus on essential task aspects without delving into the AI's internal workings [81–83]. Thus, XAI can enhance efficiency in conflict resolution tasks by reducing the cognitive demands on users. The reduction in cognitive load through explainable AI suggests that visual tools, such as HeatMap, could be broadly implemented to enhance operational efficiency in ATM. Practically, incorporating these tools could lead to fewer errors and quicker decision-making. Theoretically, this supports the cognitive load theory [84], indicating that well-designed XAI systems can serve as cognitive aids, improving user performance by simplifying complex information processing [85, 86]. This also implies that future AI development should focus on creating intuitive, easily interpretable visual aids to support high-stakes decision-making.

It is also important to recognize that the explanatory needs of controllers are not homogeneous across ATC roles. Tower, approach, and en-route controllers operate under markedly different temporal constraints and cognitive demands, which shapes how explanations should be designed and presented. Tower control requires immediate, tactical interventions in highly dynamic environments; in this setting, explanations need to be extremely concise and directly actionable, as any additional cognitive burden could compromise safety. Approach control combines both tactical sequencing and broader sector coordination, suggesting that explanations may need to balance brevity with contextual justification [87]. En-route control, by contrast, involves longer time horizons and broader situational management, making more elaborative and exploratory explanations [88], such as the HeatMap-based visualization tested here, particularly relevant. Because most experts in our sample had extensive en-route experience, our findings are most directly interpretable in this context. Nevertheless, future research should investigate how different explanation modalities may need to be adapted to role-specific operational environments, ensuring that XAI interfaces are responsive to the distinct cognitive ecologies of tower, approach, and en-route operations.

Furthermore, the type of AI explanation and user experience both influence implicit user reactions, with the HeatMap condition eliciting more positive implicit responses compared to the BlackBox condition, suggesting that visual explanations help users anchor their decision-making process. The positive response to HeatMap explanations

suggests that visual anchors are effective tools for enhancing user acceptance and trust in AI systems [89–91]. Practically, this means that incorporating visual elements that provide clear, actionable insights can improve user interaction with AI. Theoretically, this reinforces the importance of anchoring heuristics in decision-making processes, suggesting that XAI designs should incorporate these principles to facilitate more effective cognitive processing and user satisfaction. Future research should investigate other forms of visual aids that can serve similar anchoring functions across various contexts and user groups.

An additional theoretical lens relevant to these findings is the concept of we-agency, which refers to the experience of feeling a sense of control over actions performed by a partner during a collaborative task [92–94]. In the context of human-machine interactions, this sense of we-agency is often attenuated as machine-generated actions typically fall outside the human motor repertoire and are therefore less likely to be integrated into the individuals' sensorimotor predictions [95, 96]. As a result, building a sense of shared control with AI systems remains challenging. In our study, however, the positive effects of HeatMap-based XAI explanations, including reduced cognitive workload and more favorable implicit responses, suggest that visual cues may serve as cognitive scaffolds that help users simulate shared understanding or even rudimentary forms of co-agency. This effect appeared more pronounced among less experienced users, who reported a greater impact and a higher willingness to adopt the system, possibly due to greater cognitive flexibility or less anchorage to established procedural routines than experienced users [97–99]. Conversely, some experienced ATCOs preferred black-box AI, potentially reflecting a reluctance to cognitively engage with systems that disrupt established workflows [100, 101]. Taken together, these findings highlight the need for neuroadaptive or expertise-sensitive XAI systems that can dynamically tailor explanations to the user's mental model and cognitive needs. By promoting internalization of machine behaviour and aligning explanatory strategies with the user's expertise, such systems could reinforce effective we-agency, enhance trust calibration, and improve coordination fluency in human-AI teams within safety-critical environments [22, 92]. Future research should explore other forms of visual aids that can serve similar anchoring functions across various contexts and user groups.

Finally, in their comprehensive survey addressing the opacity of complex AI systems, [20] proposed a unified, user-centered definition of explainability alongside structured taxonomies for general machine learning models and deep learning approaches, highlighting critical challenges such as the interpretability-performance trade-off, the lack of standardized evaluation metrics, and the need for context-aware, user-adapted explanations. Similarly, [21] emphasized an urgent shift toward user-centric XAI systems in ATM, stressing the necessity of aligning AI outputs closely with the operational needs and cognitive models of ATCOs. Addressing these critical gaps, the current research significantly contributes to the XAI domain by empirically elucidating how visual explanations, specifically HeatMap-based XAI, influence cognitive workload, user acceptance, and intention to adopt AI decision-support tools within ATC. By differentiating user responses based on expertise, this work offers nuanced insights that inform targeted strategies for AI training and deployment tailored to diverse user groups. Moreover, by combining objective neurophysiological

metrics (EEG) with subjective assessments, this study advances methodological rigor in human-AI interaction evaluations, directly responding to recent calls for multi-modal approaches in XAI research [19, 25, 102], thereby laying practical foundations for enhancing operational efficiency, reducing cognitive fatigue, and supporting user trust in AI-supported ATC environments.

## 5 Limitations

While our findings provide promising evidence, the limited sample size constrains generalizability. Results should be considered exploratory, warranting replication with larger and more representative cohorts. Indeed, due to their professional duties and operational responsibilities, ATCOs are notoriously challenging to recruit for experimental studies. Recruitment processes typically must be initiated very early, often requiring extensive coordination with operational management. This limitation underscores the need for caution in broadly interpreting the findings and suggests that future studies should strive to secure larger and more representative samples to enhance statistical power and generalizability.

Furthermore, the exploratory and preliminary results discussed in this paper are specific to the ATM domain and may not be directly generalizable to other safety-critical domains. Each domain has unique operational requirements, user profiles, and varying levels of AI integration, all of which significantly impact the transferability of results. Even within aviation, the applicability of these findings can vary considerably. For instance, the operational requirements and explanatory needs differ substantially between en-route ATCOs, who manage long-term traffic flow over expansive airspace, and tower or approach sector ATCOs, who handle immediate, detailed tasks such as take-offs, landings, ground movements and sequencing. Given that most of our expert participants had extensive en-route experience (with some having additional tower/approach backgrounds), the results are most directly interpretable for the en-route context. However, we explicitly acknowledge that different explanation modalities may be required for tower and approach control. For example, in high-density tower operations, simpler visual cues (e.g., colour-coded alerts or minimal graphical overlays) may be preferable to full HeatMap. Such variability implies that a one-size-fits-all approach to XAI may not be effective, highlighting the necessity for tailored explanations that address the distinct demands of each role according to each task [73, 103, 104]. Indeed, recent studies further emphasize that domain-specific expertise significantly affects trust in AI systems, suggesting explanations might be particularly critical for highly experienced users [75].

Another limitation of the present work lies in the categorization of participants into only two groups: “experts” and “students.” While this dichotomy allowed for clear statistical contrasts given the limited sample size, it oversimplifies the real spectrum of expertise in ATC. In practice, expertise is a continuum shaped not only by years of experience but also by sector-specific exposure (tower, approach, en-route), the phase of professional training, and generational familiarity with novel technologies. Intermediate groups, such as recently licensed controllers in their first years of duty, or controllers who operate in specific sectors (e.g., only approach control), may

show distinct patterns of trust, workload management, and reliance on AI explanations compared to both novices and highly experienced instructors [105]. Indeed, prior research suggests that early-career professionals often rely more heavily on external aids while developing internalized strategies, whereas very experienced operators may resist automation that conflicts with established routines [106]. Capturing these nuances would provide a richer understanding of how explanatory interfaces should adapt across the full career trajectory. Future research should thus include such intermediate categories to capture more nuanced insights into how expertise modulates workload, trust, and acceptance of AI explanations across the career trajectory of controllers.

A further limitation of the present study is that only a HeatMap-based visual explanation was tested against a black-box condition. While this choice reflects the strong visual orientation of ATC work and the need for explanations that integrate directly into radar-like displays, it does not allow conclusions about other explanation modalities. Prior research in visual analytics and cognitive ergonomics suggests that textual explanations can monopolize attention and divert cognitive resources away from monitoring tasks [107], making them less suitable for real-time operational settings. By contrast, hybrid approaches, such as combining minimal text with schematic visuals, may offer richer support in training or planning scenarios, where time constraints are less acute. Future studies should therefore systematically compare different explanation modalities, including visual, textual, and hybrid formats, across ATC roles and contexts. Such comparative research would help identify when and for whom each modality is most beneficial, ultimately informing the design of flexible XAI systems that adapt explanation styles to both task demands and user preferences.

Finally, from a methodological perspective, several constraints warrant explicit acknowledgment. Electroencephalography, while providing objective neurophysiological insights, comes with inherent methodological limitations, including susceptibility to noise, potential confounding factors such as fatigue or stress, and challenges related to data interpretation due to the complexity of cognitive workload constructs. Additionally, the subjective assessment of user perceptions employed in this study was limited to only five questionnaire items, which may have restricted the breadth and depth of the captured user experiences. Another limitation is the absence of direct behavioral performance metrics such as conflict-resolution speed, accuracy, and error rates. While EEG-based neurometrics and self-reports provided complementary insights into workload and acceptance, the lack of behavioral data prevents us from confirming whether these cognitive and perceptual benefits translate into measurable improvements in operational performance. Future studies should therefore integrate behavioral logging alongside neurophysiological and subjective assessments to provide a more comprehensive evaluation of human–AI interactions. Moreover, although the present study was conducted in ENAC’s high-fidelity ACHIL en-route simulator, which offers a realistic approximation of operational conditions, simulators cannot fully reproduce the unpredictability and pressures of live operations. Thus, the transferability of these findings to real-world ATC, particularly in tower and approach sectors, remains to be established, and further research in live or shadow-mode operational environments is required.

## 6 Conclusions

This study provides critical insights into the role of visual explanations, specifically HeatMap-based XAI, in shaping cognitive workload, user acceptance, and adoption intentions within ATC environments. By addressing existing gaps in the literature concerning the empirical evaluation of XAI systems, this work demonstrated that visual explanations significantly alleviate cognitive load and enhance user acceptance (through willingness to use), underscoring the practical value of transparent, interpretable AI systems in safety-critical domains. Our findings emphasize the nuanced role of user expertise, highlighting that while visual explanations universally reduce cognitive workload and foster positive adoption intentions, explicit perceptions of AI impact are notably influenced by user experience. Less experienced users exhibited greater receptivity to AI-driven tools, suggesting that expertise and generational factors significantly shape the integration and perceived utility of novel technologies. Furthermore, this research underscores the methodological importance of employing multimodal evaluation frameworks, which integrate both objective neurophysiological measures (EEG) and subjective user feedback, thereby providing a more comprehensive understanding of human-AI interactions. Overall, this study contributes foundational knowledge for developing cognitively aligned, user-centric AI tools that effectively balance technological innovation with human needs, ultimately enhancing operational efficiency, reducing cognitive strain, and boosting trust in AI-supported ATC systems.

## Declarations

### Availability of data and materials

The datasets generated and/or analysed during the current study are available in the ARTIMATION H2020 Project ZENODO repository: <https://zenodo.org/records/7437564>; <https://zenodo.org/records/7437969>; <https://zenodo.org/records/7437777>.

### Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### Funding

This research is funded by the H2020-SESAR project “Transparent artificial intelligence and automation to air traffic management systems”, ARTIMATION, (GA No. 894238) and partially by Horizon Europe-SESAR project “Trustworthy intelligent system for remote digital tower”, TRUSTY (GA No. 101114838). The consortium would also like to express its sincere thanks to TCI: Neurophysiological Framework for Predicting Team Dynamics in High-Responsibility Operational Environments – Team-Computer Interface provided by the Sapienza University of Rome to Pietro Aricò (CUP B83C24007070005).

## Author contribution

**GC:** Conceptualization, Writing - Original Draft, Formal analysis. **AV:** Writing - Original Draft, Review & Editing. **NC:** Formal analysis, Investigation, Writing - Original Draft. **CH:** Methodology, Investigation, Resources. **AD:** Code and Data Curation. **ND:** Code and Date Curation. **AF:** Writing - Review & Editing. **MUA:** Project administration, Writing - Review & Editing. **SB (Shahina):** Project administration, Writing - Review & Editing. **SB (Shaibal):** Writing - Review & Editing. **BMSI:** Data Curation. **VR:** Data Curation, Formal analysis. **GB:** Data Curation, Formal analysis. **GDF:** Data Curation, Formal analysis. **FB:** Funding acquisition, Project administration. **PA:** Supervision, Conceptualization.

## Acknowledgements

The consortium would like to express its warm thanks to the controllers and student controllers who took part in this study.

## Appendix A Questionnaires

In this appendix are presented the questionnaires used during the experimentation with air traffic controllers. Each condition (BlackBox or HeatMap) was followed by two questionnaires: a 3-item solution agreement/comprehension questionnaire (assessing agreement with the proposed solution, perceived comprehensibility, and confidence in the solution); and a multi-item willingness-to-use/trust/acceptance questionnaire (assessing willingness to use the system in the future, trust in the system, situational awareness, acceptance, and perceived impact on work performance).

## References

- [1] Abdillah, R.E., Moenaf, H., Rasyid, L.F., Achmad, S., Sutoyo, R.: Implementation of artificial intelligence on air traffic control-a systematic literature review. In: 2024 18th International Conference on Ubiquitous Information Management and Communication (IMCOM), pp. 1–7 (2024). IEEE
- [2] Gosling, G.D.: Identification of artificial intelligence applications in air traffic control. *Transportation Research Part A: General* **21**(1), 27–38 (1987). Elsevier
- [3] Xie, Y., Pongsakornsathien, N., Gardi, A., Sabatini, R.: Explanation of machine-learning solutions in air-traffic management. *Aerospace* **8**(8), 224 (2021). MDPI
- [4] Xie, Y., Gardi, A., Liang, M., Sabatini, R.: Ai-based dynamic re-routing for dense low-altitude air traffic management. In: 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC), pp. 1–10 (2023). IEEE
- [5] Metzger, U., Parasuraman, R.: The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human factors* **43**(4), 519–528 (2001). SAGE Publications Sage CA

**Post Scenario Questionnaire:**

I agree with the proposed solution:

Yes	No
-----	----

The solution was easy to understand:

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree

I understand why this solution has been generated:

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

**Post Condition Questionnaire:****Usability**

Learn to operate the tool would be easy for me:

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

I find the tool clear and understandable:

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

I find the tool easy to use:

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

**Trust**

I felt confident when using the tool:

**Fig. A1** Questionnaire used during the experimentation (1)

- [6] Metzger, U., Parasuraman, R.: Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. In: Decision Making in Aviation, pp. 345–360 (2017). Routledge
- [7] Stefani, T., Jameel, M., Gerdes, I., Hunger, R., Bruder, C., Hoemann, E., Christensen, J.M., Girija, A.A., Köster, F., Krüger, T., *et al.*: Towards an operational design domain for safe human-ai teaming in the field of ai-based air traffic controller operations. In: 2024 AIAA DATC/IEEE 43rd Digital Avionics Systems

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

**Situational Awareness**

The tool improved my Situation Awareness of the conflicts presented:

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

**Acceptability**

I would like to use this tool in the future:

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

I like the new decision support interface:

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

**Impact on work performance**

Using this tool in my job would allow me to solve conflicts faster:

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

Using this tool in my job would increase my accuracy in solving conflicts (e.g., less mileage...):

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

Using this tool would improve my work performance:

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

Using this tool would make my work easier:

Strongly disagree	1	2	3	4	5	Strongly agree
-------------------	---	---	---	---	---	----------------

Fig. A2 Questionnaire used during the experimentation (2)

Conference (DASC), pp. 1–10 (2024). IEEE

- [8] Westin, C., Hilburn, B., Borst, C., Van Kampen, E.-J., Bång, M.: Building transparent and personalized ai support in air traffic control. In: 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), pp. 1–8 (2020). IEEE
- [9] Kirwan, B.: Human factors requirements for human-ai teaming in aviation. *Future Transportation* 5(2), 42 (2025). MDPI

- [10] Meier, J., Finke, M., Ohneiser, O., Jameel, M.: Flexible air traffic controller deployment with artificial intelligence based decision support: Literature survey and evaluation framework. Deutscher Luft-und Raumfahrtkongress 2024 (2024)
- [11] Shmelova, T., Sikirda, Y.: Artificial intelligence for evaluating the mental workload of air traffic controllers. In: Evaluating Mental Workload for Improved Workplace Performance, pp. 184–212 (2020). IGI Global Scientific Publishing
- [12] Bolić, T., Ravenhill, P.: Sesar: The past, present, and future of european air traffic management research. Engineering **7**(4), 448–451 (2021). Elsevier
- [13] Pham, D.-T., Ali, H., Fennedy, K., Hsieh, M.-H., Alam, S., Duong, V.: Human-ai hybrid paradigm for collaborative air traffic management systems. Proceedings of the SESAR Innovation Days (2024)
- [14] Ziakkas, D., Henneberry, D., Pechlivanis, K.: The role of artificial intelligence (ai) & future applications in the implementation of aviation fatigue risk management system. Human Interaction and Emerging Technologies (IHET-AI 2025): Artificial Intelligence and Future Applications **3**(3) (2025). AHFE Open Acces
- [15] Xia, Z., Chen, C.-H., Hsieh, M.-H., Ma, G., Liu, B., Yu, X., Xin, S., Dong, L.: A systematic review on human-ai hybrid systems and human factors in air traffic management. Journal of Engineering Design, 1–49 (2025). Taylor & Francis
- [16] Hernandez, C.S., Ayo, S., Panagiotakopoulos, D.: An explainable artificial intelligence (xai) framework for improving trust in automated atm tools. In: 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), pp. 1–10 (2021). IEEE
- [17] Perez-Cerrolaza, J., Abella, J., Borg, M., Donzella, C., Cerquides, J., Cazorla, F.J., Englund, C., Tauber, M., Nikolakopoulos, G., Flores, J.L.: Artificial intelligence for safety-critical systems in industrial and transportation domains: A survey. ACM Computing Surveys **56**(7), 1–40 (2024). ACM New York, NY
- [18] Wang, Y., Chung, S.H.: Artificial intelligence in safety-critical systems: a systematic review. Industrial Management & Data Systems **122**(2), 442–470 (2022). Emerald Publishing Limited
- [19] Ahmed, M.U., Barua, S., Begum, S., Jmoona, W., Cruze, R.S., Veyrie, A., Kebir, S., Hurter, C.: Role of multi-modal machine learning, explainable ai and human-ai teaming in trusted intelligent systems for remote digital towers. In: Proceedings of the 2024 7th Artificial Intelligence and Cloud Computing Conference, pp. 26–35 (2024)
- [20] Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., *et al.*: Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges

toward responsible ai. *Information fusion* **58**, 82–115 (2020). Elsevier

- [21] Degas, A., Islam, M.R., Hurter, C., Barua, S., Rahman, H., Poudel, M., Ruscio, D., Ahmed, M.U., Begum, S., Rahman, M.A., Bonelli, S., Cartocci, G., Di Flumeri, G., Borghini, G., Babiloni, F., Aricò, P.: A survey on artificial intelligence (ai) and explainable ai in air traffic management: Current trends and development with future research trajectory. *Applied Sciences* **12**(3), 1295 (2022)
- [22] Fennedy, K., Hilburn, B., Nadirsha, T.N., Alam, S., Le, K.-D., Li, H.: Do atcos need explanations, and why? towards atco-centered explainable ai for conflict resolution advisories. arXiv preprint arXiv:2505.03117 (2025)
- [23] Hurter, C., Degas, A., Guibert, A., Poyer, M., Durand, N., Veyrie, A., Ferreira, A., Cavagnetto, N., Bonelli, S., Ahmed, M., Jmoona, W., Barua, S., Begum, S., Cartocci, G., Flumeri, G.D., Borghini, G., Babiloni, F., Aricó, P.: Examining decision-making in air traffic control: Enhancing transparency and decision support through machine learning, explanation, and visualization: A case study. In: *International Conference on Agents and Artificial Intelligence*, vol. 2, pp. 622–634 (2024)
- [24] Islam, M.R., Ahmed, M.U., Barua, S., Begum, S.: A systematic review of explainable artificial intelligence in terms of different application domains and tasks. *Applied Sciences* **12**(3), 1353 (2022). MDPI
- [25] Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* **267**, 1–38 (2019). Elsevier
- [26] Saarela, M., Podgorelec, V.: Recent applications of explainable ai (xai): A systematic literature review. *Applied Sciences* **14**(19), 8884 (2024). MDPI
- [27] Hurter, C., Degas, A., Guibert, A., Durand, N., Ferreira, A., Cavagnetto, N., Islam, M.R., Barua, S., Ahmed, M.U., Begum, S., Bonelli, S., Cartocci, G., Di Flumeri, G., Borghini, G., Babiloni, F., Aricò, P.: Usage of more transparent and explainable conflict resolution algorithm: Air traffic controller feedback. *Transportation Research Procedia* **66**, 270–278 (2022)
- [28] Aricò, P., Borghini, G., Di Flumeri, G., Bonelli, S., Golfetti, A., Graziani, I., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., *et al.*: Human factors and neurophysiological metrics in air traffic control: a critical review. *IEEE reviews in biomedical engineering* **10**, 250–263 (2017). IEEE
- [29] Hilburn, B.: Cognitive complexity in air traffic control: A literature review. *EEC note* **4**(04), 1–80 (2004). Eurocontrol Brussels, Belgium
- [30] Histon, J.M., Hansman, R.J., Gottlieb, B., Kleinwaks, H., Yenson, S., Delahaye, D., Puechmorel, S.: Structural considerations and cognitive complexity in air

- traffic control. In: Proceedings. The 21st Digital Avionics Systems Conference, vol. 1, pp. 1–212 (2002). IEEE
- [31] Loft, S., Sanderson, P., Neal, A., Mooij, M.: Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human factors* **49**(3), 376–399 (2007). SAGE Publications Sage CA
- [32] Haag, F., Stingl, C., Zerfass, K., Hopf, K., Staake, T.: Overcoming anchoring bias: The potential of ai and xai-based decision support. arXiv preprint arXiv:2405.04972 (2024)
- [33] Senoner, J., Netland, T., Feuerriegel, S.: Using explainable artificial intelligence to improve process quality: evidence from semiconductor manufacturing. *Management Science* **68**(8), 5704–5723 (2022). INFORMS
- [34] Hoffman, R.R., Miller, T., Klein, G., Mueller, S.T., Clancey, W.J.: Increasing the value of xai for users: a psychological perspective. *KI-Künstliche Intelligenz* **37**(2), 237–247 (2023). Springer
- [35] Theis, S., Jentzsch, S., Deligiannaki, F., Berro, C., Raulf, A.P., Bruder, C.: Requirements for explainability and acceptance of artificial intelligence in collaborative work. In: International Conference on Human-computer Interaction, pp. 355–380 (2023). Springer
- [36] Szymanski, M., Millecamp, M., Verbert, K.: Visual, textual or hybrid: the effect of user expertise on different explanations. In: Proceedings of the 26th International Conference on Intelligent User Interfaces, pp. 109–119 (2021)
- [37] Wickens, C.D.: Multiple resources and mental workload. *Human factors* **50**(3), 449–455 (2008). SAGE Publications Sage CA
- [38] Wickens, C.D.: Situation awareness: Review of mica endsley’s 1995 articles on situation awareness theory and measurement. *Human factors* **50**(3), 397–403 (2008). SAGE Publications Sage CA
- [39] Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319–340 (1989). JSTOR
- [40] Venkatesh, V., Davis, F.D.: A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science* **46**(2), 186–204 (2000). INFORMS
- [41] Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425–478 (2003). JSTOR
- [42] Chesterman, S.: Through a glass, darkly: artificial intelligence and the problem

- of opacity. *The American Journal of Comparative Law* **69**(2), 271–294 (2021). Oxford University Press UK
- [43] Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., Babiloni, F.: Adaptive automation triggered by eeg-based mental workload index: A passive brain-computer interface application in realistic air traffic control environment. *Frontiers in Human Neuroscience* **10**, 539 (2016)
- [44] Oakley, B.: Conflict detection and resolution in air traffic control. *The Journal of Navigation* **17**(4), 433–448 (1964). Cambridge University Press
- [45] Xu, X., Rantanen, E.M.: Conflict detection in air traffic control: A task analysis, a literature review, and a need for further research. In: *Proceedings of the 12th International Symposium on Aviation Psychology*, pp. 1289–1295 (2003). Wright State University Press Dayton, OH
- [46] Durand, N., Alliot, J.-M.: Ant colony optimization for air traffic conflict resolution. In: *Atm Seminar 2009, 8th Usa/europe Air Traffic Management Research and Development Seminar* (2009)
- [47] Elsayed, S.M., Sarker, R.A., Essam, D.L.: A new genetic algorithm for solving optimization problems. *Engineering Applications of Artificial Intelligence* **27**, 57–69 (2014). Elsevier
- [48] Herrera, F., Lozano, M., Verdegay, J.L.: Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artificial intelligence review* **12**(4), 265–319 (1998). Springer
- [49] Immanuel, S.D., Chakraborty, U.K.: Genetic algorithm: An approach on optimization. In: *2019 International Conference on Communication and Electronics Systems (ICCES)*, pp. 701–708 (2019). IEEE
- [50] Willett, P.: Genetic algorithms in molecular recognition and design. *Trends in Biotechnology* **13**(12), 516–521 (1995). Elsevier
- [51] Scheepens, R., Willems, N., Wetering, H., Andrienko, G., Andrienko, N., Van Wijk, J.J.: Composite density maps for multivariate trajectories. *IEEE transactions on visualization and computer graphics* **17**(12), 2518–2527 (2011). IEEE
- [52] Asal, B., Demir, M.Ö.: Enhancing software defect prediction through explainable ai: Integrating shap and lime in a voting classifier framework. In: *2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP)*, pp. 1–7 (2024). IEEE

- [53] Roshinta, T.A., Gábor, S.: A comparative study of lime and shap for enhancing trustworthiness and efficiency in explainable ai systems. In: 2024 IEEE International Conference on Computing (ICOCO), pp. 134–139 (2024). IEEE
- [54] Ullah, I., Rios, A., Gala, V., McKeever, S.: Explaining deep learning models for tabular data using layer-wise relevance propagation. *Applied Sciences* **12**(1), 136 (2021)
- [55] Sciaraffa, N., Di Flumeri, G., Germano, D., Giorgi, A., Di Florio, A., Borghini, G., Vozzi, A., Ronca, V., Babiloni, F., Aricò, P.: Evaluation of a new lightweight eeg technology for translational applications of passive brain-computer interfaces. *Frontiers in Human Neuroscience* **16**, 901387 (2022). Frontiers Media SA
- [56] Sciaraffa, N., Di Flumeri, G., Germano, D., Giorgi, A., Di Florio, A., Borghini, G., Vozzi, A., Ronca, V., Varga, R., Gasteren, M., *et al.*: Validation of a light eeg-based measure for real-time stress monitoring during realistic driving. *Brain sciences* **12**(3), 304 (2022). MDPI
- [57] Aricò, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., Colosimo, A., Babiloni, F.: Passive bci in operational environments: Insights, recent advances, and future trends. *IEEE Transactions on Biomedical Engineering* **64**(7), 1431–1436 (2017)
- [58] Borghini, G., Aricò, P., Astolfi, L., Toppi, J., Cincotti, F., Mattia, D., Cherubino, P., Vecchiato, G., Maglione, A.G., Graziani, I., *et al.*: Frontal eeg theta changes assess the training improvements of novices in flight simulation tasks. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 6619–6622 (2013). IEEE
- [59] Gevins, A., Smith, M.E., McEvoy, L., Yu, D.: High-resolution eeg mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral cortex (New York, NY: 1991)* **7**(4), 374–385 (1997)
- [60] Guan, K., Zhang, Z., Liu, T., Niu, H.: Cross-task mental workload recognition based on eeg tensor representation and transfer learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **31**, 2632–2639 (2023). IEEE
- [61] Sciaraffa, N., Aricò, P., Borghini, G., Flumeri, G.D., Florio, A.D., Babiloni, F.: On the use of machine learning for eeg-based workload assessment: Algorithms comparison in a realistic task. In: *International Symposium on Human Mental Workload: Models and Applications*, pp. 170–185 (2019). Springer
- [62] Ekman, P.E., Davidson, R.J.: *The Nature of Emotion: Fundamental Questions.*, (1994). Oxford University Press
- [63] Harmon-Jones, E., Harmon-Jones, C., Fearn, M., Sigelman, J.D., Johnson, P.:

- Left frontal cortical activation and spreading of alternatives: tests of the action-based model of dissonance. *Journal of personality and social psychology* **94**(1), 1 (2008). American Psychological Association
- [64] Hugdahl, K., Nordby, H., Kvale, G.: Conditional learning and brain asymmetry: Empirical data and a theoretical framework. *Learning and Individual Differences* **1**(4), 385–405 (1989). Elsevier
- [65] Cartocci, G., Modica, E., Rossi, D., Cherubino, P., Maglione, A.G., Colosimo, A., Trettel, A., Mancini, M., Babiloni, F.: Neurophysiological measures of the perception of antismoking public service announcements among young population. *Frontiers in human neuroscience* **12**, 231 (2018). Frontiers Media SA
- [66] Cartocci, G., Rossi, D., Modica, E., Maglione, A.G., Martinez Levy, A.C., Cherubino, P., Canettieri, P., Combi, M., Rea, R., Gatti, L., *et al.*: Neurodante: poetry mentally engages more experts but moves more non-experts, and for both the cerebral approach tendency goes hand in hand with the cerebral effort. *Brain Sciences* **11**(3), 281 (2021). MDPI
- [67] Ronca, V., Di Flumeri, G., Giorgi, A., Vozzi, A., Capotorto, R., Germano, D., Sciaraffa, N., Borghini, G., Babiloni, F., Aricò, P.: o-clean: A novel multi-stage algorithm for the ocular artifacts' correction from eeg data in out-of-the-lab applications. *Journal of Neural Engineering* **21**(5), 056023 (2024). IOP Publishing
- [68] Delorme, A., Makeig, S.: Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods* **134**(1), 9–21 (2004). Elsevier
- [69] Klimesch, W.: Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews* **29**(2-3), 169–195 (1999). Elsevier
- [70] Aricò, P., Reynal, M., Imbert, J.-P., Hurter, C., Borghini, G., Di Flumeri, G., Sciaraffa, N., Di Florio, A., Terenzi, M., Ferreira, A., *et al.*: Human-machine interaction assessment by neurophysiological measures: a study on professional air traffic controllers. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4619–4622 (2018). IEEE
- [71] Cartocci, G., Maglione, A.G., Rossi, D., Modica, E., Borghini, G., Malerba, P., Piccioni, L.O., Babiloni, F.: Alpha and theta eeg variations as indices of listening effort to be implemented in neurofeedback among cochlear implant users. In: *International Workshop on Symbiotic Interaction*, pp. 30–41 (2017). Springer
- [72] Cartocci, G., Caratù, M., Modica, E., Maglione, A.G., Rossi, D., Cherubino, P., Babiloni, F.: Electroencephalographic, heart rate, and galvanic skin response

- assessment for an advertising perception study: application to antismoking public service announcements. *Journal of visualized experiments: JoVE* (126), 55872 (2017)
- [73] Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., *et al.*: Guidelines for human-ai interaction. In: *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems*, pp. 1–13 (2019)
- [74] Cocchioni, M., Bonelli, S., Westin, C., Borst, C., Bang, M., Hilburn, B.: Learning for air traffic management: guidelines for future ai systems. In: *Journal of Physics: Conference Series*, vol. 2526, p. 012105 (2023). IOP Publishing
- [75] Bayer, S., Gimpel, H., Markgraf, M.: The role of domain expertise in trusting and following explainable ai decision support systems. *Journal of Decision Systems* **32**(1), 110–138 (2022). Taylor & Francis
- [76] Newendorp, A.K., Gilbert, S.B., Dorneich, M.C.: User’s ai expertise (aix): Measuring one’s mental model of an ai system. In: *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, pp. 1–8 (2024). IEEE
- [77] Halawi, L., Miller Ed D, M.D., Holley Ph D, S.J., *et al.*: Artificial intelligence in aviation: A path analysis. *Journal of Aviation/Aerospace Education & Research* **33**(4), 10 (2024)
- [78] Bui, L., Pezzola, M., Bandara, D.: How do ai explanations affect human-ai trust? In: *International Conference on Human-Computer Interaction*, pp. 175–183 (2023). Springer
- [79] Maehigashi, A., Fukuchi, Y., Yamada, S.: Adjusting amount of ai explanation for visual tasks. In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–7 (2024)
- [80] Chen, C., Liao, M., Sundar, S.S.: When to explain? exploring the effects of explanation timing on user perceptions and trust in ai systems. In: *Proceedings of the Second International Symposium on Trustworthy Autonomous Systems*, pp. 1–17 (2024)
- [81] Bhagya, J., Thomas, J., Raj, E.D.: Exploring explainability and transparency in deep neural networks: A comparative approach. In: *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 664–669 (2023). IEEE
- [82] Hudon, A., Demazure, T., Karran, A., Léger, P.-M., Sénécal, S.: Explainable artificial intelligence (xai): how the visualization of ai predictions affects user cognitive load and confidence. In: *NeuroIS Retreat*, pp. 237–246 (2021). Springer

- [83] Mesinovic, M., Watkinson, P., Zhu, T.: Explainable ai for clinical risk prediction: a survey of concepts, methods, and modalities. arXiv preprint arXiv:2308.08407 (2023)
- [84] Kahneman, D.: The human side of decision making. *J Invest Consult* **13**, 9–14 (2012)
- [85] Chen, Z., Luo, Y., Sra, M.: Engaging with ai: How interface design shapes human-ai collaboration in high-stakes decision-making. arXiv preprint arXiv:2501.16627 (2025)
- [86] Soltan, M.A.A., Abdellatif, M.M.: Psychology of visual communication in ai-based decision aid tools for engineering applications. In: 2024 International Conference on Decision Aid Sciences and Applications (DASA), pp. 1–5 (2024). IEEE
- [87] Miller, M., Holley, S.: Air traffic controller resource management: An approach for reducing cognitive loading and increasing situational awareness. In: International Conference on Applied Human Factors and Ergonomics, pp. 535–542 (2021). Springer
- [88] Inoue, S., Furuta, K., Kanno, T., Aoyama, H., Nakata, K.: Cognitive process modelling of team cooperative work in en route air traffic control. *IFAC Proceedings Volumes* **43**(13), 19–24 (2010)
- [89] Cho, I., Wesslen, R., Karduni, A., Santhanam, S., Shaikh, S., Dou, W.: The anchoring effect in decision-making with visual analytics. In: 2017 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 116–126 (2017). IEEE
- [90] Karran, A.J., Demazure, T., Hudon, A., Senecal, S., Léger, P.-M.: Designing for confidence: The impact of visualizing artificial intelligence decisions. *Frontiers in neuroscience* **16**, 883385 (2022). Frontiers Media SA
- [91] Morrison, K., Shin, D., Holstein, K., Perer, A.: Evaluating the impact of human explanation strategies on human-ai visual decision-making. *Proceedings of the ACM on Human-Computer Interaction* **7**(CSCW1), 1–37 (2023). ACM New York, NY, USA
- [92] Haggard, P.: Sense of agency in the human brain. *Nature Reviews Neuroscience* **18**(4), 196–207 (2017). Nature Publishing Group UK London
- [93] Pacherie, E.: The sense of control and the sense of agency. *Psyche* **13**(1), 1–30 (2007)
- [94] Pacherie, E.: How does it feel to act together? *Phenomenology and the cognitive sciences* **13**(1), 25–46 (2014). Springer

- [95] Sahai, A., Pacherie, E., Grynszpan, O., Berberian, B.: Predictive mechanisms are not involved the same way during human-human vs. human-machine interactions: A review. *Frontiers in neurorobotics* **11**, 52 (2017). Frontiers Media SA
- [96] Sahai, A., Desantis, A., Grynszpan, O., Pacherie, E., Berberian, B.: Action co-representation and the sense of agency during a joint simon task: Comparing human and machine co-agents. *Consciousness and cognition* **67**, 44–55 (2019). Elsevier
- [97] Chi, M.T.: Two approaches to the study of experts' characteristics. *The Cambridge handbook of expertise and expert performance*, 21–30 (2006)
- [98] Endsley, M.R.: Situation awareness misconceptions and misunderstandings. *Journal of cognitive Engineering and Decision making* **9**(1), 4–32 (2015). Sage Publications Sage CA
- [99] Krems, J.F.: Cognitive flexibility and complex problem solving. In: *Complex Problem Solving*, pp. 201–218 (2014). Psychology Press
- [100] Parasuraman, R., Manzey, D.H.: Complacency and bias in human use of automation: An attentional integration. *Human factors* **52**(3), 381–410 (2010). Sage Publications Sage CA
- [101] Woods, D., Dekker, S.: Anticipating the effects of technological change: A new era of dynamics for human factors. *Theoretical issues in ergonomics science* **1**(3), 272–282 (2000). Taylor & Francis
- [102] Graziani, M., Dutkiewicz, L., Calvaresi, D., Amorim, J.P., Yordanova, K., Vered, M., Nair, R., Abreu, P.H., Blanke, T., Pulignano, V., *et al.*: A global taxonomy of interpretable ai: unifying the terminology for the technical and social sciences. *Artificial intelligence review* **56**(4), 3473–3504 (2023). Springer
- [103] Abdul, A., Vermeulen, J., Wang, D., Lim, B.Y., Kankanhalli, M.: Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–18 (2018)
- [104] Nguyen, T., Canossa, A., Zhu, J.: How human-centered explainable ai interface are designed and evaluated: A systematic survey. *arXiv preprint arXiv:2403.14496* (2024)
- [105] Timotic, D., Netjasov, F., Cicevic, S.: Investigation of relationship between trust in automation and human personality traits among air traffic controllers. *ICRAT* (2020)
- [106] Mirchi, T., Vu, K.-P., Miles, J., Sturre, L., Curtis, S., Strybel, T.Z.: Air traffic

controller trust in automation in nextgen. *Procedia Manufacturing* **3**, 2482–2488 (2015)

[107] Bertin, J.: *Semiology of Graphics*. University of Wisconsin press, ??? (1983)

ARTICLE IN PRESS