Neuroergonomics and Cognitive Engineering, Vol. 102, 2023, 83–92
https://doi.org/10.54941/ahfe1003008

Guidelines for Artificial Intelligence in Air Traffic Management: A Contribution to EASA Strategy

Matteo Cocchioni[1], Stefano Bonelli[1], Carl Westin[2], Ana Ferreira[1], and Nicola Cavagnetto[1]

[1]Deep Blue SRL, 00185, Rome, Italy
[2]Linköping University, 581 83, Linköping, Sweden

ABSTRACT

Artificial intelligence has the potential to improve air traffic management through the consistent use of machine learning. AI can bring benefits to air traffic controllers in terms of workload, situational awareness, trust, and thus operational efficiency and safety. However, human problem-solving strategies can potentially collide with AI and lead to misunderstandings and a decrease in user acceptance of air traffic control systems. The proposed paper focuses on the design of the ML system, in particular providing insights and guidelines derived from results of recent field studies as they addressed the impacts of conformance and transparency on controller behaviour and survey responses. Several guidelines were distilled based on empirical insights obtained from experiments, feedback from controllers and workshop results. The guidelines are divided into different categories: ML/AI design, Personalization, Transparency, and HCI. The proposed paper also describes a contribution to a different use case to test the generalizability of the guidelines themselves, as well as a recent update in the explainability framework developed by a regulatory authority.

Keywords: Human factors, Artificial intelligence, Explainability, Air traffic control

INTRODUCTION

The MAHALO and ARTIMATION projects, which were funded by SESAR 3 Joint Undertaking program recently finished two years of technical research on the effects of AI and Machine Learning on human performance in en-route air traffic control. ARTIMATION investigated how much transparent algorithms can help Air Traffic Controllers (ATCo) to better understand and accept solutions proposed by the ML system in the context of conflict resolution. The teams behind MAHALO at the same time, posed a crucial question: should AI be designed to mimic the strategies and style of air traffic controllers (termed "strategic conformance"), or should it be transparent and easily understood by controllers? (Westin et al., 2016) (Figure 1). These two approaches have distinct implications for building trust and understanding of automation, when it is challenging to develop systems that are easy for humans to interpret and interact with. Overcoming such difficulties is seen as essential to building meaningful collaboration between humans and

		TRANSPARENCY	
		Low	**High**
CONFORMA	**Low**	*"It's doing a strange thing, and I don't understand why…"*	*"It's doing a strange thing, but I understand why…"*
	High	*"It's doing the right thing, but I don't understand why…"*	*"It's doing the right thing, and I understand why…"*

Figure 1: Strategic conformance and transparency can vary independently.

autonomous agents (Chen et al., 2014; Endsley et al., 2017; Lyons et al., 2021).

The MAHALO project began by creating a hybrid Machine Learning system for conflict detection and resolution, along with a real-time simulation platform and experimental user interface. After several development trials, the project culminated in two field studies conducted in two European countries, involving a total of 36 ATCos. During these studies, controllers' behaviour was recorded in a pre-test phase and used to train the strategic conformance ML system. The main experiment trials then manipulated the strategic conformance of the ML models (as either personalized, group average, or optimized) and the transparency of the conflict resolution advisories (as either a basic vector depiction, an enhanced graphical diagram, or a diagram-plus-text presentation). The results (Westin et al., 2022) were measured by objective performance and behavioural data, as well as self-reported workload and survey responses. The results revealed a significant impact of strategic conformance on controllers' response to advisories, with controllers responding more positively to advisories that matched their preferred separation distance. No main effects of advisory transparency were found, but transparency did interact with strategic conformance.

ARTIMATION's experiment (ARTIMATION Deliverable 6.2 Validation report. 2023) consisted in a low fidelity human in the loop simulation with 21 participants (11 professional ATCOs and 10 ATCO students). The duration was one hour of conflict resolution tasks using three explanation conditions: (1) Black Box, where only the selected solution is presented, (2) Heat Map, where a corpus of potential solution is displayed thanks to a density map, (3) Story Telling, where data driven storytelling technique is applied to convey the explication of the proposed solution. The data was collected through debriefings at the end of the session, over-the-shoulder observations, questionnaires, and neurophysiological measurements.

GUIDELINES FOR FUTURE AI SYSTEMS IN ATC

From the experiments of the two projects, feedback from controllers, and workshop results, multiple guidelines have then been derived. They address the integration of AI/ML systems into air control tasks and their effects on controller acceptance, workload, and understanding of the system, and are divided into four main categories:

Table 1. Guidelines for future AI systems in ATC: ML/AI design.

N.	Guideline
1.1	Future AI systems for ATC should investigate which ML models are best suited for balancing individual preferences and optimization approaches.
1.2	Data collection should be performed over a long time period, ensuring a large data pool, in order to facilitate ML generated personalized and non-biased outputs.
1.3	The selection of parameters should be carefully considered because it can introduce bias in the AI solutions/support provided. Meaning that the parameters defining an optimal solution could not be the best ones to take into consideration, making the proposal optimal in abstract contexts, but less optimal in a specific operational scenario.

1. ML/AI design
2. Personalization
3. Transparency
4. Human-Computer Interaction (HCI)

Below is an in-depth written description about them.

Machine Learning/Artificial Intelligence Design

ML/AI techniques can provide benefits in solving complex traffic problems by considering multiple goals, however, they are more appropriate for pre-tactical phases like multi-sector planning and airspace management, which have a higher degree of uncertainty, rather than tactical operations where controllers are faced with solving ad-hoc sector perturbations with lower uncertainty. Additionally, when humans need to work with computerized agents that make decisions, the system should be consistent and predictable, as advocated by the Human-Centered Automation (HCA) perspective. ML solutions are generally based on probabilities and therefore less predictable than traditional deterministic CD&R algorithms.

A good practice would be to design ML/AI systems at different levels of complexity so that a fallback option is available when the highest complexity levels are infeasible with limited training data, as the required amount of data for ML/AI systems is often underestimated.

Personalization

The MAHALO and ARTIMATION projects have established that personalization can improve the acceptability and ease of use of ML/AI-based advisory systems in ATC (e.g. faster response times). Personalization can be seen as a means of addressing some of the challenges associated with the integration of ML/AI techniques in ATC from the perspective of human-machine collaboration. However, a prerequisite for personalization in decision-making is the presence of sufficient intra-controller consistency and inter-controller variability in terms of actions/clearances, which requires a large dataset to determine. MAHALO has demonstrated that there is a sufficient basis for personalization in ATC decision-making. However, it should be noted that

Table 2. Guidelines for future AI systems in ATC: personalisation.

N.	Guideline
2.1	The development of future personalized AI systems for ATC requires end users' involvement in model development. This way the model may capture what operators consider important for problem solving in the target task.
2.2	ML models require large amounts of individual data for stable understanding of problem-solving preferences. Synthetic data can be used to supplement other sources in model development.
2.3	A suitable individual preference parameter for personalizing CD&R systems in conflict resolution choices is target separation distance.
2.4	Future ATC systems that are more personalized may reduce the need for them being transparent.
2.5	AI Transparency should seek personalization to potentially achieve higher acceptance from users. It has been verified that some valid solutions have been rejected simply because they did not correspond to the user decision-making strategy, which might ultimately increase their workload in assessing the solution.
2.6	Future ATC systems should acknowledge and embrace in the design that controllers differ in their conflict resolution preferences.
2.7	Future ATC systems should consider personalized applications when possible (i.e., taking into account a safety risk assessment).
2.8	Decision support systems capable of providing advisories/recommendations on actions should do so before the operator has made a decision on how to act (note that this can be before the action is implemented).
2.9	What aspects of a system that should be personalized should be driven by the operator's individual preferences in working and problem solving, and in what regards the operator is consistent over time.
2.10	In future AI systems for air traffic control, the goal should not be to solely replicate human behaviour and decision-making, but to optimize solutions while taking into account individual operator preferences. The system should also be able to provide reasoning for its solution if it differs from the individual's preference.

personalization using supervised ML/AI techniques that model and mimic an individual human controller may result in suboptimal performance of the ATC system. It is therefore recommended that the performance of personalized advisory systems is evaluated against target Key Performance Indicators, regardless of sufficient intra-controller consistency and inter-controller variability. Having a ML algorithm that learns 'Principles' and 'Control strategies' context based on a single ATCO or to a wider category based on ATCO control strategies type would be a way forward to improve AI support based on ATCOs feedback.

Transparency

Similar to personalization, transparency can enhance the acceptability and understanding of ML-based advisory systems in ATC. The MAHALO project adopted an ecological approach to operationalize ML transparency, by emphasizing interpretable (visual) representations rather than the explainable ML

Table 3. Guidelines for future AI systems in ATC: transparency.

N.	Guideline
3.1	AI transparency should be considered for tasks or operational phases that are not timely constrained.
3.2	AI systems for ATC should focus on increasing transparency when human and system approach differ or human understanding is poor.
3.3	Ecological interface design approaches can be used to increase the transparency of presented CD&R advisories by providing information on the constraints and solution possibilities affecting the control problem.
3.4	Future AI systems in ATC should explore using transparency to enhance design. However, increased transparency may not always lead to increased acceptance or agreement with its advisories.
3.5	Increased transparency improves understanding of the system, its output, and the situation. This allows the operator to better assess if the system's behaviour or advisories are suitable for the problem, and how they align with the operator's preferences.
3.6	Transparency should be customized for individual users to foster human-AI dialogue. To do this, AI must be able to understand the user's needs and preferences.
3.7	Transparency should be considered to support humans in building trust in AI tools.

models commonly found in XAI research. This approach was well-received by controllers, as it contextualizes ML solutions within the problem at hand (i.e., traffic conflict). The same representation also served as a decision support tool, allowing controllers to formulate their own solutions and/or adjust the advised ML solution. This raises the question of what an operational controller might want and need to understand about the automated system and to what extent. For example, an ATCo might not require a deep understanding of ML neural networks at the level of a ML system developer. Furthermore, transparency needs are also affected by workload demands - in time-critical situations, an ATCo generally prefers to receive any workable solution and may not want to devote valuable cognitive resources to understanding that solution by investigating layers of information. In such cases, information on the resulting aircraft separation targeted by the advisory would be sufficient. Given that transparency needs are likely context-dependent and sensitive to operator preferences, we recommend an adaptive approach that allows ATCos to contextualize machine decisions and decide upon what they wish to see and when. It's worth noting that such an adaptive approach can be seen as another form of personalization, one that focuses on the preferred information one wishes to see. ARTIMATION in its experiment explored the transparency from the visual explanation point of view. In general, experts were less optimistic about the conflict resolution visualisation in terms of performance improvement. Higher transparency was considered more useful for less timely critical or tasks or operational phases in which the ATCOs are subject to lower risk of cognitive workload, like planning tasks.

Furthermore, it is important to note that also from the *Human Computer Interaction* point of view there are recommendations. Interaction flexibility is vital for ATCo engagement, as human involvement and final responsibility for the safety of operations is crucial in the ATC system. The MAHALO project demonstrated one approach to facilitating interaction by integrating it into existing controller tools. Through a conventional clearance menu, ATCos were able not only to accept, adjust, or change machine advisories, but also reject them and work with other aircraft than the one receiving the advisory. This flexibility was generally well-received by ATCos as it gave them the ability to influence the system in any way they preferred. We believe that the benefits of this flexibility outweigh any potential performance decrements that may occur when ATCos change an optimal advisory into a suboptimal one. It's worth noting that providing flexibility in interaction can also be considered a form of personalization.

First Exploitation Activities

These guidelines have begun to contribute to the definition of other aviation use cases involving the implementation of systems based on Artificial Intelligence and Machine Learning, even if not directly related to Conflict Detection & Resolution tasks. Such an exploitation process can be seen as a further attempt to validate the goodness of these guidelines for different applications, demonstrating that they are generalizable. In this context, within the HAIKU project, it has been developed a use case of Urban Air Mobility (UAM) with the potential to precisely demonstrate this generalizability, through the development of a Digital Assistant.

In the near future, the use of Urban Air Mobility is expected to increase, requiring new solutions for managing and regulating traffic in cities. HAIKU envisions the use of digital assistants to support human coordinators in managing UAM traffic and ensuring safety and efficiency. In this scenario, the definition and division of roles are crucial to create the best team between human and artificial operators. It is good practice to conduct such an assessment at the interaction and system design stage to avoid developing a product that will later prove to be either biased, not accepted by end users, or unsafe.

The guidelines previously discussed contributed to the definition of the *Human AI Teaming* aspects of this use case, where a UAM Coordinator is expected to play a key role in providing strategic and tactical services to UAM operators and stakeholders, while a *Digital Assistant for UAM Coordinator* (DUC) will handle standard tasks and reduce workload for the human coordinator. It is then planned that the DUC will also support the human coordinator in both normal operations and emergency situations, while the UAM Coordinator and DUC will work together to safely monitor U-space and provide services. The UAM Coordinator will be able to coordinate ground and air activities, respond to emergencies, and manage flow of aircraft, while the DUC will handle standard tasks, allowing the UAM Coordinator to focus on high-level decision making. Furthermore, in emergency situations, DUC and UAM Coordinator will provide assistance, inform stakeholders, and adhere to procedures, such as configuring safety boundaries and prioritizing flights.

It is important to mention that the development of such a use case is still in early stages, but it is expected that the guidelines will continue to help in the next steps of more technical implementation of the digital assistant.

Contribution to EASA

The European Union Aviation Safety Agency (EASA) established an internal task force on AI in October 2018 with the aim of creating a roadmap (EASA Artificial Intelligence Roadmap 1.0) that outlines key opportunities and challenges of AI in aviation and how they may impact the Agency in terms of organization, processes, and regulations. The roadmap seeks to establish EASA's vision on AI development in the aviation domain and foster interaction with stakeholders. It served so far as a dynamic document to be revised and improved as the Agency gains experience and stakeholders provide input. Thus, it served as a basis for the Agency's ongoing work in this area.

As stated in that document, "to ensure the evaluation of AI/ML applications' conformity to EU ethical principles, guidance on a comprehensive AI trustworthiness analysis needed to be developed by the Agency". Such analysis encompasses the seven principles (Figure 2) outlined in the EU ethical guidelines (accountability, technical robustness and safety, oversight, privacy and data governance, non-discrimination and fairness, transparency, and societal and environmental well-being), while the guidance provides applicants with a checklist of essential considerations for using AI/ML applications in their product designs. Completing this trustworthiness analysis and addressing any ethical concerns it may raise is a necessary step before proceeding with the evaluation of an AI/ML application's acceptability within a project. Outcomes from this analysis then should consider other 3 building blocks, namely Leaning Assurance, AI Explainability and AI safety Risk Mitigation.

Within this framework, the guidelines described in the present paper, in addition to helping to frame different aviation use cases than just Conflict Detection & Resolution scenarios, were presented precisely to a team of EASA experts, aiming at receiving input and feedback for the continued improvement of the framework itself.

What emerged was that the guidelines offered indications of greater granularity within the "AI Explainability" building block, not at a technical level, i.e., algorithmic, but at a much more operational one and concerning the activities of frontline operators for whom the personalization, transparency, and human-machine interaction aspects are crucial. EASA itself recently released an expansion of the Trustworthy AI building blocks (EASA First usable guidance for Level 1 machine learning applications - Issue 01) (Figure 3), in which it introduces a clearer differentiation between *Development Explainability*, to be achieved at an algorithmic phase, and *Operational Explainability*, which instead must be pursued within the front lines of operators. More generally, this concept fits into the broader impact of what Human Factors can bring to Artificial Intelligence and Machine Learning in safety critical

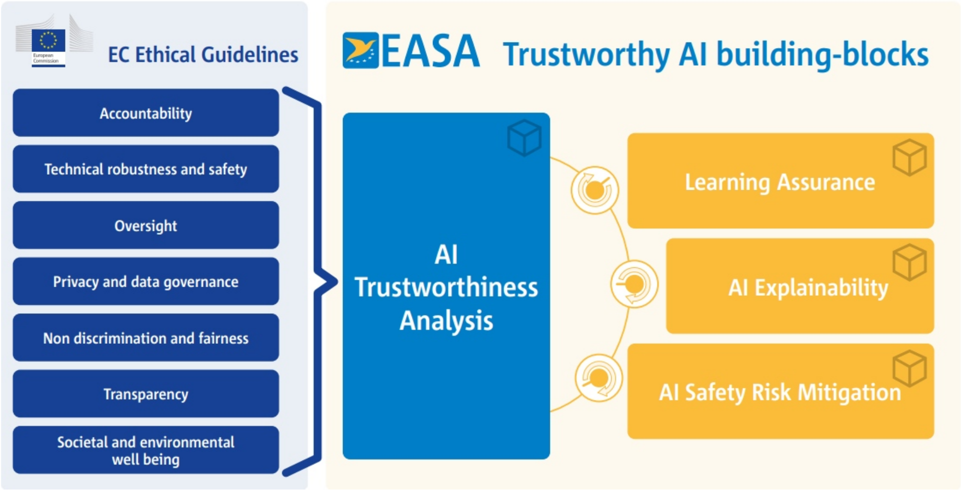

Figure 2: Relationship between AI roadmap building blocks and AI trustworthiness.

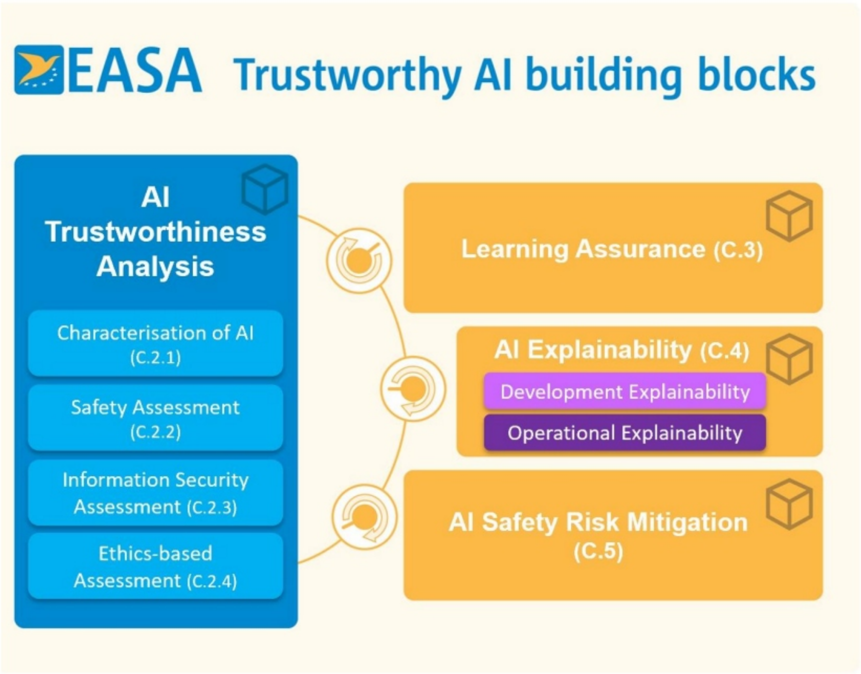

Figure 3: Expansion of trustworthy AI blocks.

organizations. There, concepts like Human AI Teaming will become more prominent soon and, again, the constructs of personalization, transparency, and human-machine interaction will become more decisive.

DISCUSSION

The MAHALO and ARTIMATION projects offer guidelines on incorporating strategic conformance and transparency in AI solutions for CD&R and safety-critical systems. They concluded several design guidelines across different categories, with the highest benefits seen in highly conformal automation with low transparency. As long as the system is effective, transparency is not necessarily required.

The traditional approach to system design is to create a uniform system that all operators must conform to, which is suitable for standard procedures and situations. However, this is less effective in time-critical and safety-critical situations, where the definition of optimal solutions is subjective and depends on the operator's ability to manage the tasks. Research showed that a person's response to conflict resolution advice is partly based on how closely it aligns with their preferences. System's objectives should not only be to conform to individual preferences but to find the best, most efficient, and safest solution when it can be defined. When that is ambiguous, or human acceptance and trust are essential, systems should consider individual preferences. Transparency should also be tailored to foster communication between humans and AI. When there is a disagreement between the system and human, transparency, in the form of an explanation of the system's behaviour or reasoning, is critical to building trust. The dialogue can be initiated by either the system sensing disagreement or the operator seeking clarification.

Future AI systems should accommodate individual preferences while generating resolution advisories and providing explanations, but further research is needed to understand the implications of implementing strategic conformal automation and its impact on user responses. For this reason, in the next years, it is important to test the generalizability and acceptability of these principles to other use case studies and practical applications, as well as involve the regulatory bodies in the certification process. Improvements have been accomplished in the last few years, but in a fast-changing environment, many are yet to come.

ACKNOWLEDGMENT

This work was developed under the HAIKU project that received funding from Horizon Europe program, grant agreement ID 101075332. This work was also developed under MAHALO and ARTIMATION projects that received funding from SESAR 3 Joint Undertaking grant agreements respectively No. 892970 and No. 699381. The SESAR 3 JU receives support from the European Union's H2020 research and innovation programme.

The authors wish to acknowledge the air traffic controllers for their precious time during the numerous experiments' runs, and for sharing their useful insight on the developed concept.

REFERENCES

ARTIMATION Deliverable 6.2 Validation report. 2023. https://www.artimation.eu/wp-content/uploads/2023/01/D6.2Validation-report_Artimation_v_00.02.00.pdf

EASA Artificial Intelligence Roadmap 1.0, 7th of February 2020, https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-roadmap-10-published

EASA First usable guidance for Level 1 machine learning applications -Issue 01, https://www.easa.europa.eu/en/downloads/134357/en

J. Y. C. Chen and M. J. Barnes, “Human-agent teaming for multirobot control: A review of human factors issues,” IEEE Trans. Human–Mach. Syst., vol. 44, no. 1, pp. 13–29, 2014.

M. R. Endsley, “From here to autonomy: Lessons learned from human–automation research,” Hum. Factors, vol. 59, no. 1, pp. 5–27, 2017.

J. B. Lyons, K. Sycara, M. Lewis, and C. A, “Human–autonomy teaming: Definitions, debates, and directions,” Front. Psychol., vol. 12, no. 589585, 2021.

Westin, C., Borst, C., & Hilburn, B. (2016). Strategic Conformance: Overcoming acceptance issues of decision aiding automation? IEEE Trans. Human–Mach. Syst., 46(1), 41–52. doi: 10.1109/THMS.2015.2482480.

C. Westin, B. Hilburn, C. Borst, E-J. van Kampen, T. Nunes, S. Boonsong, M. Cocchioni, and S. Bonelli. (2022), Personalizerd and transparent AI support for ATC conflict detection and resolution: an empirical study. In 12th SID, Dec. 5-8, 2022, Budapest, Hungary.