**Transparent and Personalised AI Support in Air Traffic Control: First Empirical Results**

Bonelli, S.1, Borst, C.2 , Brambati, F.1, Cocchioni, M.1, Hilburn, B.3 & Monteiro-Nunes, T.2\*\*\*

\*\*\* authors are ordered alphabetically by surname

1 Deep Blue SRL, Rome, Italy

2 Deft University of Technology, Delft, the Netherlands

3 Center for Human Performance Research CHPR BV, Voorburg, the Netherlands

**Introduction**

The air traffic management (ATM) community is committed to greater use of automation to accommodate predicted traffic growth. In particular, Artificial Intelligence (AI) is seen as a key enabler of future ATM systems. The ATM community, in turn, is actively engaged in research to help guide the design of AI in future ATM.

Against this backdrop, the SESAR-sponsored MAHALO (Modern ATM via Human-Automation Learning Optimisation) project is exploring a simple but profound question: in the emerging age of Machine Learning (ML), should we be developing AI that matches air traffic controller behavior (i.e., is strategically *conformal*), or AI that is understandable (i.e., *transparent*) to the controller [1]? Further, what tradeoffs exist, in terms of controller trust, acceptance, and performance? The latter issue is currently receiving a great deal of research attention The growing Explainable Artificial Intelligence (XAI) movement recognizes that system transparency and understandability are key to the successful development of ‘third wave’ AI [2]. However, relatively little attention has been paid to the issue of strategic conformance. Figure 1 presents the hypothetical tradeoff between AI conformance and transparency.

|  |  |  |
| --- | --- | --- |
|  |  | **TRANSPARENCY** |
|  |  | Low | High |
| **CONFORMANCE** | Low | **Stupid AI:***“It’s doing a strange thing, and I don’t understand why…”* | **Peculiar AI:***“It’s doing a strange thing, but I understand why…”* |
| High | **Confusing AI:** *“It’s doing the right thing, but I don’t understand why…”* | **Perfect AI:***“It’s doing the right thing, and I understand why…”* |

**Figure 1. AI conformance vs transparency.**

MAHALO set out to assess how strategic conformance and transparency of a machine learning decision support system for conflict detection and resolution affect air traffic controllers’ understanding, trust, acceptance, and workload of its advice and performance in solving conflicts, and how these factors (conformance and transparency) interact.

To answer these questions, the MAHALO project has:

* Developed an individually-tuned hybrid ML system comprised of Supervised Learning (based on a deep learning neural network approach) and a Reinforcement Learning (based on Deep Q-Learning from Demonstrations, DQfD approach) models to process conflict detection- and resolution tasks, respectively. Pre-test data were used to create personalized solutions, based on Supervised Learning;
* Coupled this to an enhanced enroute CD&R prototype display to present machine rationale with regards to ML output;
* Conducted control-in-the-loop simulations to assess the relative impact of ML *conformance, transparency,* andtraffic *complexity*, on controller understanding, trust, acceptance, workload, and performance.

**Methods**

Two controller-in-the-loop simulations were conducted in Dec 2021 / Jan 2022 (n=20) and Mar-Apr 2022 (n=16). Each of these two simulations consisted of a *Conformance Pre-Test* and a follow-up *Main Experiment.* In the former, controllers actively managed a series of short air traffic scenarios. Conflict scenarios and controller responses / resolutions were captured, and used to train the ML system, which was then active in the follow-up Main Experiment phase, as described below.

In the Main Experiment, two independent variables were varied within controllers, as follows:

Conformance— was defined as the match between individual and ML advisories. The three levels of conformance were: (a) *Individual* (i.e., an exact replay of that controller’s pre-test solution, made unrecognizable through heading and callsign changes); (b) *Group average* (in which the ML system was trained on data from the entire group of controllers, and (c*) Optimal* (in which the RL system presented an optimized solution, per a reward function and set of optimization rules).

Transparency— was defined as the degree to which the system explains its underlying rationale. The three levels of transparency were: (a) *Baseline*, in which system-generated solutions were presented with a resolution heading vector, but no further information; (b) *Domain transparency*, in which ‘NO-GO” polygons depicted flightpath constraints, based on proximate aircraft; and (c) *Agent transparency*, in which additional text messages presented the rationale for the solution. The three levels of transparency are shown in Figure 2.



**Figure 2. Three levels of transparency.**

In the main experiment, the following dependent variables were collected or derived:

* Acceptance rate – controllers were free to accept, modify, or reject and replace any given advisory. Data from simulation 1 indicate that outright rejection was rare. Acceptance was categorized as *accept*, *nudge* (i.e., a slight change in heading value), or *adjust* (a change in value and direction).
* Agreement—a subjective measure of agreement, once per advisory. Notice that this can, and often does, differ from acceptance (e.g. time pressure might force one to accept a solution for which agreement is low);
* Solution bias—the tendency to adopt a more liberal / conservative trajectory (i.e., a smaller / larger closest point of approach (CPA)) could be captured as a bias in solution strategies.
* Subjective workload—rated onscreen at the end of each scenario;
* Subjective understanding—at the end of each scenario
* Response times to conflict and solution
* Post session questionnaire responses—Thirteen Likert (1-6) agreement scale items on such issues as trust, perceived safety, accuracy, timing, and usability of solutions and the user interface.

**Results**

As of this writing, the first of the two simulations has been conducted, and preliminary analysis has been carried out. Analysis of the full data set can only be completed after simulation 2 finishes in late April, 2022. Thus, the proposed paper and presentation (in-person) would include an analysis of that full data set, along with lessons learnt.

**References**

[1] MAHALO (2021). Concept Report, deliverable D2.2, SESAR Joint Undertaking (SJU) grant number 892970.

[2] Launchbury, J. (2017). A DARPA Perspective on Artificial Intelligence. Online at: <https://www.darpa.mil/about-us/darpa-perspective-on-ai>