Just Culture in the Era of Digitalization: How Artificial Intelligence is Expected to Influence Just Culture in the Air Traffic Management System?

by Stathis Malakis and Marc Baumgartner***

Abstract

While the Air Traffic Management system is fast transiting towards the vision of a Digital European Sky and Artificial Intelligence is seen as the key enabler, the questions concerning Just Culture have developed into: how Artificial Intelligence is expected to influence Just Culture and do we need to rewrite the Just Culture playbook? In this article we provide an initial attempt to answer these questions. We enumerate the limitations that are expected to be influential and identify two layers of concerns that prompt the rewriting of the Just Culture playbook.

1. Introduction

It is not far too long that computers were seen as foolproof machines that processed numerical inputs into numerical outputs and whose calculations were never wrong. Nowadays, digital machines ranging from smartphones and tablets to personal computers and data warehouses are dealing with humanlike tasks that go beyond basic number crunching and enter the realm of higher cognitive processes such as information analysis, pattern recognition, predictive insights, and decision-making. These are achieved via the ubiquitous application of Artificial Intelligence (AI) and Machine Learning (ML). AI and ML have already proved themselves to be viable technologies, and their applications in many domains are increasing every day. This trend of collecting and learning from data is expected to continue even stronger in the near future[1]. Analysis of data allows us to both understand the process that underlies the past data and also predict the behavior of the process in the future. A system that is in a changing environment should have the ability to learn; otherwise, we would hardly call it intelligent. ML is not just a database or programming problem; it is also a requirement for AI that enables learning.

From disruptive events to pandemics, political unrest, war conflicts and climate change, winning the future depends on adaptation. To survive and thrive, organizations must maintain a competitive advantage and enable an ability to win in a way that doesn't just withstand changes but embrace them to generate new strategic possibilities. The imperative for change is increasingly the creation of an adaptable organization—one that can thrive in the digital economy. An adaptive organization in the 21st century is archetypically digitally powered, leading many organizations to pursue digital transformation. Air Navigation Service Providers (ANSPs), which are the building blocks of the Air Traffic Management (ATM) system are not an exception to this rule. Sustained adaptability is a constant call for ANSPs and refers to the ability to continue to adapt to changing environments, stakeholders, demands, contexts, and constraints within the wider aviation system.

* Stathis Malakis PhD, Air Traffic Controller, Hellenic Aviation Service Provider & IFATCA Joint Cognitive Human Machine Systems Group. This article is written in a personal capacity and any views expressed are those of the author and do not necessarily reflect the views of any organization the author is affiliated with. Email: stathis.malakis@gmail.com

** Marc Baumgartner, Swiss Air Traffic Controller, IFATCA SESAR and EASA coordinator. This article is written in a personal capacity and any views expressed are those of the author and do not necessarily reflect the views of any organization the author is affiliated with. Email: sesar.coord@ifatca.org

[1] Alpaydın, E. (2021). Machine learning: the new AI. Cambridge, MA: MIT Press.

In recent years, there has been a realization that AI/ML is introduced in the safety critical domains building upon:

- advances in capacity to collect and store massive amounts of data;
- significant increases in computing power, and
- development of increasingly powerful algorithms and architectures.

The main advantages of AI/ML are the following:

- rapid identification of patterns in complex real-world data that humans and conventional computer assisted analyses struggle to identify;
- real-time support in decision making, and optimization of multi-parameter problems.

The ATM system is a leading example of a cognitively complex safety critical system where AI/ML applications are introduced. ATM systems comprise several airspace sectors and aerodromes with varying air traffic flows that interact in complex ways and evolve dynamically. As safety critical situations arise (e.g. separation minima between aircraft are violated), Air Traffic Controllers are applying clearances, instructions and information in order to keep safe separations between aircraft[2]. Controllers are expert decision makers who employ cognitive strategies developed over years of operational experience, recurrent training, and shear accumulation of ATC systems knowledge. Quite often improvise in situ to meet the challenges of traffic imposed by novel events, unfortunate actions and shortcomings of the work system.

The ATC is a complex safety critical system with the following work characteristics[3]:

- rapidly escalating situations: the transition from normal to abnormal situations can be rapid (e.g., an aircraft experiencing a decompression can initiate a rapid descend of 6,000 ft per minute affecting the safety of other aircraft, without any prior notice);
- time pressure: available time for decision-making and coordination is severely constrained (e.g., in a loss of separation scenario, the conflict geometry must be detected and resolved within a few seconds);
- severe error consequences: errors can lead to disasters when compensating mechanisms (e.g., automation safety nets) are not present or act in destabilizing ways;
- complex, multi-component decisions: Air Traffic Controllers may have several degrees of freedom (e.g., change the flight level, the speed and the route of aircraft) but their decisions may be in conflict with other goals;
- conflicting/shifting goals: the goals of Safe, Orderly and Expeditious flow as imposed by the International Civil Aviation Authority[4] are conflicting and consequences may cascade from system design down to the tactical level of day-to-day operations. For example, an orderly flow of air traffic may be a safe flow but not an expeditious one. Another example refers to following noise abatement procedures in the selection of the runway in use when weather conditions dictate the change of runway.

[2] ICAO, (2016). Air traffic management. Procedures for air navigation services, Document 4444. 16 ed. Montreal: International Civil Aviation Organization.

[3] Malakis, S., Kontogiannis, T., and Kirwan, B. (2010). Managing emergencies and abnormal situations in air traffic control (Part I): Taskwork strat**egies.** Applied Ergonomics, 41, 620–627.

[4] ICAO, (2018). Air Traffic Services: ANNEX 11 to the Convention of International Civil Aviation. 15th edition. Montreal: International Civil Aviation Organization.

In the next years AI/ML systems are expected to support -more than now- Air Traffic Controllers in performing their safety critical tasks within the ATM ecosystem. Furthermore, the widespread introduction of AI/ML is expected to create a new ATM environment, which will be tightly coupled, more complex to manage, and with pressing needs for:

- minimization of delays;
- reduce the cognitive workload of the Air Traffic Controllers and thus permitting a higher throughput of aircraft in an en-route sector, approach sector and aerodrome tower;
- accommodating a diverse array of autonomous aircraft;
- operating smoothly in adverse weather conditions;
- smoothing out 4D aircraft trajectories, and minimizing environmental impact.

European Union[5] is envisioning a Digital European Sky[6] and an irrevocable shift to low and ultimately no-emission mobility. With this goal in mind, ATM and aviation will evolve into an integrated digital ecosystem characterized by distributed data services. This is planned to be accomplished mostly by leveraging digital technologies to transform the aviation sector. The aim is to deliver a fully scalable ATM system for manned and unmanned aviation that is even safer than today's, and, based on higher air ground integration. While the essence of ATM is and will always remain, to ensure the safe and orderly execution of all flights it needs to do so in the most environmentally friendly and cost-efficient way.

In the years to come ATM infrastructure will become more data-intensive. If we consider the ATM comprised of three layers, then current architecture is characterized by:

- airspace layer: limited capacity, poor scalability, fixed routes, fixed national air-space structures;
- air traffic service layer: limited automation, low level of information sharing;
- physical layer: fragmented ATM infrastructure.

Digitalization and especially the introduction of AI/ML, which lies at its core, is expected to enable:

- dynamic & cross FIR airspace configuration & management free routes high resilience at the airspace layer;
- automation support & virtualization Scalable capacity at the air traffic service layer;
- integrated and rationalized ATM infrastructure at the physical layer.

The ATM provides an essential service for aviation and AI/ML are considered the key enablers to overcome current shortcomings and provide the necessary abilities to adapt to the increasing demands of normal operations and disturbances/crisis for the ANSPs. It is envisioned that ATM practitioners will be able to design and eventually operate a system that is smarter and safer, by constantly analyzing, gaining insights, and learning from all aspects of the ATM ecosystem by utilizing AI/ML, deep learning algorithms and big data analytics.

New and emerging AI/ML capabilities are required for the future ATM and U-space environment to provide the necessary levels of performance beyond current limits. Full scale implementation of ATM virtualization that will allow the

[5] SESAR JU, (2020). Strategic Research and Innovation Agenda - Digital European Sky

[6] SESAR JU, (2020). EUROPEAN ATM MASTER PLAN- Digitalising Europe's Aviation Infra-structure.

complete decoupling of ATM service provision from the physical location of the personnel and equipment is highly dependent upon digitalization and most importantly to state of the art AI/ML algorithms.

2. Safety and Just Culture in the ATM system

Commercial Aviation counts amongst the safest means of transport per distance travelled. This success is at-tributed to the philosophy of accident and incident investigation which was developed in the early days of aviation. This has provided a rich source of understanding and learning which has led to the development of im-proved technological, procedural, organizational facets of the aviation industry as well as human factors. From the very onset of commercial aviation, a large part of the success in increasing safety has been attributed to lessons learned from investigations and hazard identifications. However, it should be recognized that many other factors have contributed to increased safety records in aviation, such as:

- continuous oversight by national regulatory agencies;
- rigorous training of front-line personnel (e.g., Air Traffic Controllers, flight crews and maintenance engineer) for their certification;
- use of airworthiness specifications;
- regulatory imposed backup and redundant safety systems;
- failure containment design principles (e.g., standard design precautions to minimize aircraft hazards in the event of engine rotor failures);
- use of Human Factors Engineering principles in design of aircraft and ATM systems;
- controlled experimentation of systems, components, and procedures (e.g., use of Level D flight simulators, use of wind tunnels to explore the aerodynamics of new designs, etc.);
- simulator training for Air Traffic Controllers;
- institutionalizing evidence-based training;
- operation of mandatory and voluntary reporting systems;
- promotion of Just Culture principles;
- independent industry, agencies, and user associations' efforts to increase safety (e.g., Flight Safety Foundation, pilots' associations, IFATCA, IATA, CANSO, AIRBUS, Boeing);
- circulation of safety knowledge in the form of narratives of safety occurrences and stories of near miss events within communities of practice (Air Traffic Controllers, flight crews, engineers, ramp agents, flight dispatchers).

In recent years, the understanding has evolved so that any incident or occurrence should be reported – voluntary or otherwise, to further increase safety of commercial aviation through the disclosure of incidents and occurrences. Mandatory and voluntary reporting schemes have been developed and are part of national and international law. Whereas accident and incident investigation shall not be used to apportion blame there is, however, a risk that the information disclosed by the reporter will be used in a criminal proceedings. The notion of Just Culture has been created as a counterbalance to keep intact the safety information flow.

James Reason is attributed as the first to coin the term 'Just Culture' and describes it as a part of a larger safety culture[7]. Reason also identified key limitations of no-blame reporting cultures. As a solution to this, the idea of a Just Culture was introduced. He introduced the language of Just Culture, foundational ideas, and principles that ground the notion, and most notably that of gross negligence. Envisaging safety culture to have five components, one of which is Just Culture, Reason argues that a just culture is not the same as a no-blame culture but is a culture where individuals are accountable for their willful misconduct or gross negligence. Reason defined Just Culture as an atmosphere of trust in which people are encouraged (even rewarded) for providing essential safety-related information, but in which they are fully aware about where the demarcation line between acceptable and unacceptable behavior is. An effective reporting culture is therefore underpinned by a 'Just Culture', in which there exists a line of acceptable practice. To this, Reason introduced an algorithm, later to be known as the substitution test, by which an individual practitioners' actions, could be examined, and tested against characteristics of the event and occurrence. Subsequently, Dekker provided further avenues by successions of books[8] and journal papers[9] of what later would become the widely accepted view by which aviation organizations and the industry conceived Just Culture.

It is also recognized that Safety Science will also need to evolve to cope with the safety challenges posed by the introduction of AI/ML. Currently, safety assurance frameworks, are not adapted to AI/ML and new ones are being developed[10].

It has long been recognized that ATM safety, is based upon open and insightful safety information *"flowing"* through the *"information veins"* of the system. The Just Culture concept as conceived by James Reason, developed by Sidney Dekker and adopted by organizations[11], is fundamental to keep the safety information flowing from the operations rooms (i.e. the sharp end of the ATM system) to the executive team (i.e. the blunt end) by institutionalizing that ATM front line operators are able, allowed and willing to share safety information by reporting incidents and other safety-related issues, and when there is a commitment to act on what is shared in order to learn and make things better.

Just Culture is defined by the European Union as[12]: *"a culture in which front-line operators and others are not punished for actions, omissions or decisions taken by them which are commensurate with their experience and training, but where gross negligence, willful violations and destructive acts are not tolerated".*

Before proceeding any further it is stressed that *"gross negligence"*, *"willful violations"* and *"destructive acts"* are legal and not human factors terms.

The concept of Just Culture essentially addresses the mutual recognition of two key functions, aviation safety and the administration of justice and represents the fundamental recognition that both would benefit from a carefully established equilibrium, moving away from fears of criminalization, balancing and satisfying the interests of the two unique and basically not compatible domains. Just Culture does not mean complete protection of front-line operators in the event of aviation incidents and accidents. Particularly, it does not offer protection in case of gross negligence, willful misconduct and/or destructive acts, severe and serious disregard of an obvious risk and/or profound failure of professional responsibility[13].

[7] Reason, J. (1997). Managing the Risks of Organizational Accidents. Aldershot: Ashgate.

[8] Dekker, S. (2012). Just Culture, Balancing Safety and Accountability. Second edition, Ashgate.

[9] Dekker, S. W. A. (2009). Just Culture: Who draws the line? Cognition, Technology and Work, 11(3), 177-185.

[10] EASA, 2020. Artificial Intelligence Roadmap 1.0. European Aviation Safety Agency.

[11] EUROCONTROL, (2018). Model for a Policy Regarding Criminal Investigation and Prosecution of Aviation and Railway Incidents and Accidents. Brussels: Eurocontrol.

[12] REGULATION (EU) No 376/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 3 April 2014 on the reporting, analysis and follow-up of occurrences in civil aviation, amending Regulation (EU) No 996/2010 of the European Parliament and of the Council and repealing Directive 2003/42/EC of the European Parliament and of the Council and Commission Regulations (EC) No 1321/2007 and (EC) No 1330/2007.

[13] HELVETICA, (2022). Just Culture Manual for ATCO, ANSE & ATSEP: Behavior after an incident and further proceedings. V2.0, October 2022. HelvetiCA (Swiss Controllers' Association).

Keeping the Just Culture equilibrium at a balance is based upon:

- the notions of acceptable and unacceptable behaviors, and
- the concept of the *"honest mistake"*.

And this is where AI/ML inherent limitations comes into play.

3. AI/ML and Just Culture

From the state-of-the-art multimodal large language models that are currently available to general public down to the dedicated AI/ML systems that support non-critical tasks these systems come with known limitations. For instance, for a multimodal large language model, known limitations may, such as social biases, hallucinations, and adversarial prompts[14]. Before the introduction of AI/ML into the ATM system it was difficult but possible to draw the red line between *"gross negligence"*, *"willful violations"* and *"destructive acts"* on the one side and *"honest mistakes"* on the other side.

To implement an AI/ML project in the Operations Rooms there are several multi-level challenges that must be addressed[15], such as:

- explainability issues of the AI/ML models. If an Air Traffic Controller is presented an AI/ML system that portends to explain how it works, how do we know whether the explanation works, and the user has achieved a pragmatic understanding of the AI/ML? This question entails some key concepts of measurement such as explanation goodness and trust[16];
- function allocation issues. Avoiding leftover strategy where the Air Traffic Controllers are expected to assume control when AI/ML fails;
- development validation and harmonization of the integration of AI/ML technologies in the whole system, among all users (Air Traffic Controllers, flight crews, aerodrome operators and Network Manager);
- keeping the Air Traffic Controllers 'in the loop' and situationally aware and able to intervene;
- disruption of established patterns in coordinated activity between Air Traffic Controllers' and between Air Traffic Controllers' and flight crews;
- disruption of established patterns of resilience;
- increasing instances of AI/ML induced surprises and clumsiness;
- demand for the development of new mental models, how the AI system works, how it fails, why it fails, and how to adapt[17].

[14] GPT-4 (Generative Pre-trained Transformer 4) created by OpenAI. https://openai.com/product/gpt-4

[15] Malakis, S. Baumgartner, M. Berzina, N. Laursen, T. Smoker, A. Poti, A. Fabris, G. (2022). Challenges from the Introduction of Artificial Intelligence in the European Air Traf-fic Management System, IFAC-PapersOnLine, Volume 55(29), 1-6

[16] Hoffman, R., Mueller, S., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance. Frontiers in Computer Science, 5. http://doi.org/10.3389/fcomp.2023.1096257. Retrieved from: https://digitalcommons.mtu.edu/michigantech-p/16886.

[17] Borders, J., Klein, G., and Besuijen, R. (2019). An operational account of mental models: A pilot study. Proceedings of the 2019 International Conference on Naturalistic Decision Making, San Francisco, CA.

Apart of them there always exist the in-build technical ones due to the nature of algorithms, data and statistics used, such as:

- avoiding data leakage between training validation and testing data sets;
- bias management. Undesirable bias should be identified, evaluated, and when possible, eliminated to contribute to data representativeness:
  - bias introduced by any sampling which could be applied to the data;
  - bias introduced when performing data cleaning or removal of presupposed outliers;
  - recall bias introduced during data annotation or data labelling;
  - bias introduced by adversarial attack resulting in data poisoning.
- capturing singularities: singularities in data refer to the presence of discontinuities or more generally speaking non-linearities characterized by steep variations of various frequency. The intensity, density and complexity of these singularities are the main obstacles to an AI/ML model accuracy;
- consistency checks against the Operational Design Domain (ODD);
- data labelling: labelled data highlights data properties, characteristics, or classifications that can be analyzed for patterns that help predict the target. Data labelling is the process of detecting and tagging target data;
- dimensionality reduction: this step aims to reduce the number of input variables by projecting input data into a lower-dimensional feature space;
- explainability of AI/ML models;
- feature selection: a method that discards the uninformative features and keeps only those that are informative; it is another method for dimensionality reduction;
- generalization: how well a model trained on a training set performs on new data unseen during training;
- hyperparameters tuning: hyper-parameters are identified during the building of the model (training agnostic) like (but not limited to) number of hidden nodes and layers, input features, learning rate, activation function in neural networks (e.g., for neural networks: number of layers, number of neurons in each layer, and their connections, selection of the activation functions in each layer, learning rate);
- identifying missing data;
- limit checks (e.g., Range limits, min., and max. values for the parameter);
- normalization and standardization (scaling): the aim is to get most inputs in the range of 0.0 to 1.0 or -1.0 to +1.0 or something similar;
- overfitting;
- removing bad data (e.g., Garbage characters or error codes);
- representativeness: a dataset is representative when it is complete, and the distribution of its key characteristics is similar to the intended space of the ODD of the targeted application;
- selection of the training stopping criterion(criteria) for ML models;
- split test vs cross validation.

For instance, state of art algorithms for AI/ML systems such as neural networks are essentially *"black boxes"* in terms of explainability. Arguably, the best-known disadvantage of neural networks is their "*black box*" nature. Simply put, you don't know how or why the neural network came up with a certain output given a certain input. In other words, they are tremendously successful in providing accurate predictions based on historical data, but no-one can understand why.

So, consider an Air Traffic Controller in the operations room who is provided a rather "odd' suggestion for a course of action from an AI/ML digital assistant that employs neural networks. If something goes wrong who is to blame?

With the introduction of AI/ML, as responsibilities for the execution of tasks are progressively delegated to technology, liability will also tend to shift from human operators to the organizations that designed and developed the technology, defined its context and uses, and are responsible for its deployment, integration, and maintenance. Such a shift implies not only that a different target will be moved on the first line for liability attribution (from human operators to technology manufacturer/users), but also that different grounds will be taken into account: from the assessment of human negligence in carrying out his/her duties, to the assessment of the defectiveness of a technology in carrying out its function. It is evident that with the introduction of AI/ML they are triggered by different conditions, have different defenses and different rules for the evaluation of evidence and the burden of proof.

This is the first level of concerns we face for Just Culture in the AI/ML era.

The second level refers to the training of Air Traffic Controllers. The definition of Just Culture speaks of *"…actions, omissions or decisions taken by them which are commensurate with their experience and training…"* but right now, Air Traffic Controllers do not receive any formal training to AI/ML and especially to the state-of-the-art algorithms such as neural network and their limitations. So, do we have to train the controllers on AI/ML and at what level? Do we have to train controllers into understanding bias-variance trade-offs, explainability issues, data validation, feature engineering, hyper-parameters selection, overfitting, limitations of data driven models and all that stuff from AI/ML before providing them with digital assistants in the OPS room?

This is the second layer of concerns.

For both layers the answers are difficult and we don't pretend we have them. Current playbooks for Just Culture cannot give definitive answers to these questions.

4. Conclusions

Changes in the ATM domain are of permanent nature and challenges of research, development, and transition to introduce these changes are a daily life for ANSPs and their staff. Be it Air Traffic Controllers, technicians, engineers, managers, and decision makers. New Technologies leading digitalization, including AI and ML are finding their ways into the ATM working environment. Whereas lot of expectation is linked to a so-called technology hype introduction of new technology will have to follow the path of introducing new technological component into a running ATM system. Linked to the regulatory and certification challenges, a lot of the modern technology will have to be interwoven into the existing architecture. This will create new challenges, surprises not only at the operational level but also – as we argue in this paper – in safety critical concepts, such as Just Culture.

The introduction of AI/ML can be so transformative as it was the RADAR in the ATC back in the 50s. Since then, it became the standard tool for the Air Traffic Controllers. Controllers now take an extensive formal and on-the-job training in the fundamentals of the radar systems theory and the data processing systems.

We don't know yet how radical this transformation will be, but we need to influence it to the right direction. Is this something that needs to change in terms of Just Culture?

The answer is yes.

We argue that the introduction of AI/ML in essence clouds the drawing of a red line between "*gross negligence*", "*willful violations*" and "*destructive acts*" on the one side and "honest mistakes" on the other side (Figure 1).

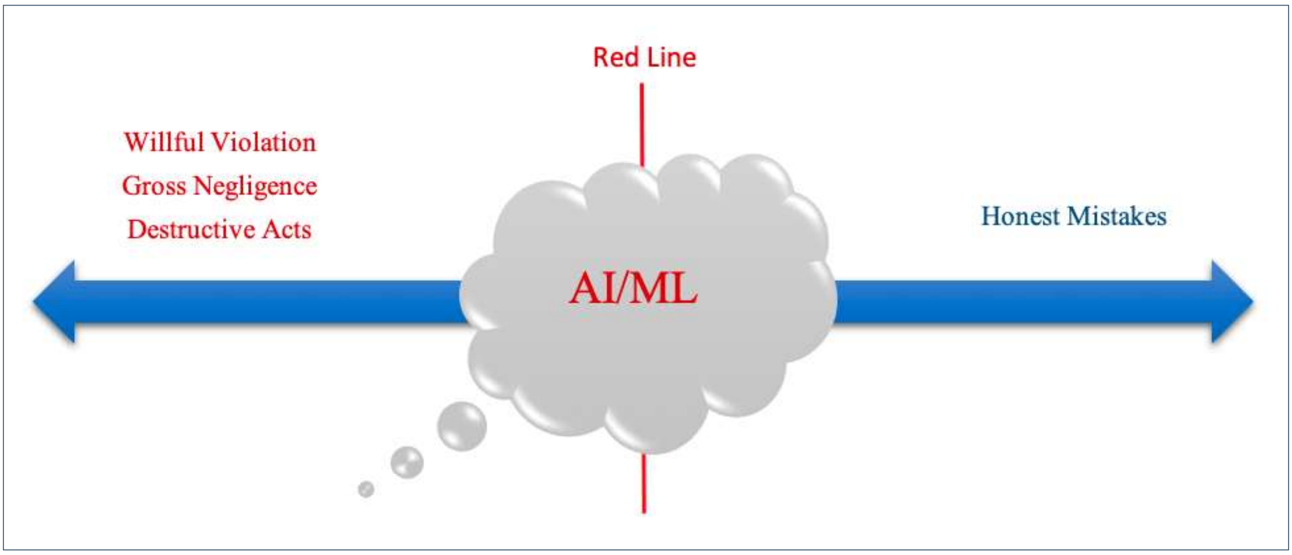

Figure 1 AI/ML clouds the drawing of a red line between "gross negligence", "willful violations" and "destructive acts" on the one side and "honest mistakes" on the other side

We need to redefine Just Culture and rewrite its playbook in the era of digitalization.