AIR POWER

Journal of Air Power and Space Studies

Vol. 18 No. 2 • Summer 2023
(April-June)

Contributors

Mr Abhishek Saxena • Mr Anubhav S. Goswami • Ms Neha Mishra
• Air Marshal Sukhchain Singh • Mr Shaurya Dhakate • Lieutenant Colonel Sten Arve
• Mr Tanuj Pandey and Wing Commander Swaim Prakash Singh

CENTRE FOR AIR POWER STUDIES, NEW DELHI

EXPLAINABLE AI FOR AIR WAR OPERATIONS

SUKHCHAIN SINGH

> We need the ability to not only have high performance models, but also to understand when we cannot trust those models.
>
> –Alexander Amini

During the USA-led invasion of Iraq, on March 22, 2003, US troops fired a Patriot interceptor missile on an assumed Iraqi anti-radiation missile which later proved to be a UK Tornado fighter jet ZG710. The Royal Air Force (RAF) investigation concluded that the error was due to the missile's computer algorithms. These were related to how the missile classified targets, rules for engagement and firing the missiles, autonomous operation of missile batteries, and technical and procedural factors, like the fighter not singing "friend or foe" at that instant.[1] It's an eye-opener into how Artificial Intelligence (AI)-enabled systems or automated tools on the battlefield will

Air Marshal **Sukhchain Singh** AVSM, VSM (Retd) was commissioned in the Aeronautical Engineering (Electronics) Branch of the Indian Air Force (IAF) in July 1979 and was the Air Officer-in-Charge Maintenance (AOM) at Air Headquarters, New Delhi, before his retirement in October 2015. He is a Gold Medallist in BE (Hons), Electronics and Communication from the Regional Engineering College (now NIT), Kurukshetra, in the year 1979 and M Tech from IIT, Delhi, in Integrated Electronics and Circuits. He is an alumnus of the Defence Services Staff College, Wellington, and, MBA in Operations Research from the Indira Gandhi National Open University (IGNOU).

1. Kelsey Atherton, "Understanding the Errors Introduced by Military AI Applications", May 6, 2022, at https://www.brookings.edu/techstream/understanding-the-errors-introduced-by-military-ai-applications/. Accessed on January 2, 2023.

Artificial Intelligence (AI) is no longer in the hypothetical field and is now deployed in tangible real-world military applications of targeting decisions of military weapons. To obviate errors, the human in the loop does prevent accidental or incorrect shootdowns.

perform in the fog of war. The incorporation of AI in military applications comes with immense risk. In combat scenarios, from the movement of autonomous air vehicles to identifying tanks on a battlefield, AI computers are expected to provide quick and accurate decisions even under errors of various hues, be they human or technology-centric. A clear understanding of how AI will fail is, therefore, important during the evaluation, trials and induction of such systems.[2]

ERRORS IN AI

Artificial Intelligence (AI) is no longer in the hypothetical field and is now deployed in tangible real-world military applications of targeting decisions of military weapons. To obviate errors, the human in the loop does prevent accidental or incorrect shootdowns which necessarily requires a fine balance of trust between human and autonomous machine decision-making.[3] However, the engineering psychologist John Hawley, involved in the U.S. Army's study of the 2003 friendly fire incidents, brought out in his 2017 report that "humans are very poor at meeting the monitoring and intervention demands imposed by supervisory control." Therefore, *a-priori*, when AI is incorporated into weapons, the sensors, and information displays should include an awareness of errors and present that information in a suitable way without adding to the reasoning load of the person using the machine.[4]

One of the principles is "traceability," and the Pentagon is investing in testing, evaluation, validation, and verification methods for AI. The development of testing and "explain-ability" tools for military AI applications represents one of the key challenges for the technology. This work is ongoing

2. Ibid.
3. Ibid.
4. Ibid.

at the Joint Artificial Intelligence Centre, which, in February 2020, awarded contracts to 79 vendors worth up to $15 million apiece to develop testing and evaluation technology.[5]

The Raksha Mantri (RM) launched 75 AI products/technologies during the first-ever 'AI in Defence' symposium and exhibition in New Delhi on July 11, 2022. It has been termed as a revolutionary step in the development of humanity. Its book form is a compendium of some stupendous products and stellar achievements in AI-based technologies in defence by individuals, companies and institutions towards catapulting the Indian defence industry into a major global force.[6] The introduction of autonomy in weapon systems, in ISR (Intelligence, Surveillance and Reconnaissance), and data management, can be a huge asset in stopping terrorism, installing counter-terrorism measures and protecting soldiers.[7] It is now imperative that a very elaborate mechanism be put in place for the evaluation and testing of these AI systems with repeatability, traceability and explainabilty before deployment in the field and combat units.

The introduction of autonomy in weapon systems, in ISR (Intelligence, Surveillance and Reconnaissance), and data management, can be a huge asset in stopping terrorism, installing counter-terrorism measures and protecting soldiers.

EXPLAINABLE AI (XAI) MADE SIMPLE

The more discerned reader is referred to a masterpiece for a thorough understanding of explainable AI: *Interpretable Machine Learning, A Guide for Making Black Box Models Explainable*, by Christoph Molnar[8].

5. Ibid.
6. Artificial Intelligence in Defence, GOI, Ministry of Defence (Department of Defence Production), at https://www.ddpmod.gov.in/sites/default/files/ai.pdf. Accessed on January 2, 2023
7. Ibid.
8. Christoph Molnar, *Interpretable Machine Learning, A Guide for Making Black Box Models Explainable* (Lean Publishing). This book is for sale at http://leanpub.com/interpretable-machine-learning. at https://christophm.github.io/interpretable-ml-book/. Accessed on January 2, 2023.

AI algorithms are "black boxes" that take inputs and provide outputs with no way to understand their inner workings. Many AI algorithms use deep learning, in which algorithms learn to identify patterns based on enormous training data. Deep learning is a neural network approach that imitates the brain. Similar to our thought processes, it is difficult to determine how a deep learning algorithm has arrived at a decision. It can be useful to subdivide XAI into three categories based on data, predictions and algorithms. These are:

- **Explainable Data:** What data went into training a model? Why was that data chosen? How was fairness assessed? Was any effort made to remove bias?
- **Explainable Predictions:** What features of a model were activated or used to reach a particular output?
- **Explainable Algorithms:** What are the individual layers that make up the model, and how do they lead to the output or prediction?

There are two current approaches to explainability. One based on proxy modelling, which is a different type of model such as a decision tree that is used to approximate the actual model. Because it's an approximation, it can differ from the true model results. Second, based on design for interpretability. These are models which are designed so that their outputs are easy to explain. This approach runs the risk of reducing the predictive power or overall accuracy of a model.[9]

TOOLS

Explainable AI frameworks are tools which generate reports about how the model works and try to explain their working. Each year, the number of available XAI tools (developed by both academics and industry practitioners) grows, resulting in more options than ever for those interested in using

9. Mike McNamara, "Explainable AI: What is it? How Does it Work? And What Role Does Data Play?" February 22, 2022, at https://www.netapp.com/blog/explainable-ai/. Accessed on January 3, 2023.

them. There were more than 150 XAI-related tools published between 2015 and 2022 alone. I will briefly explain about six such explainable AI frameworks.

SHAP stands for SHapley Additive exPlanations. It can be used for explaining various types of models like simple machine learning algorithms like linear regression, logistic regression, tree-based models, and also more complex models like deep learning models for image classification and image captioning and various Natural Language Processing (NLP) tasks like sentiment analysis, translation, and text summarisation. It is a model agnostic method to explain the models based on game theory's SHapley values. It explains how the different features affect the output or the contribution they have in the outcome of the model.

LIME stands for Local Interpretable Model-agnostic Explanations. It is similar to SHAP but is faster in terms of computation. The output of LIME is a list of explanations, reflecting the contribution of each feature to the prediction of a data sample. LIME is able to explain any black box classifier, with two or more classes. All it requires is that the classifier implements a function that takes in raw text or a numpy array and outputs a probability for each class. Support for the Machine Learning (ML) frameworks scikit-learn classifier is built-in.

ELI5 is a Python package which helps to debug machine learning classifiers and explain their predictions. It has support for many ML frameworks like scikit-learn, Keras, XGBoost, LightGBM, and CatBoost.

There are two main ways to look at a classification or a regression model:

1. Inspect model parameters and try to figure out how the model works globally.
2. Inspect an individual prediction of a model, try to figure out why the model makes the decision it makes.

The What If Tool (WIT) has been developed by Google to understand the working of ML trained models. Using WIT, you can test performance

For practitioners to successfully select a useful XAI tool, they must be able to assess the tool's ability to provide domain-appropriate explainability, understand its limitations, and have a clear idea about how it can be used to produce meaningful information about ML model predictions.

in hypothetical situations, analyse the importance of different data features, and visualise model behaviour across multiple models and subsets of input data, and for different ML fairness metrics. The What-If Tool is available as an extension in Jupyter, Collaboratory and Cloud AI Platform notebooks. It can be used for different tasks like binary classification, multi-class classification and regression. It can be used with various types of data like Tabular, Image and Text data. It can be used along with SHAP and LIME. It can also be used with Tensor Board.

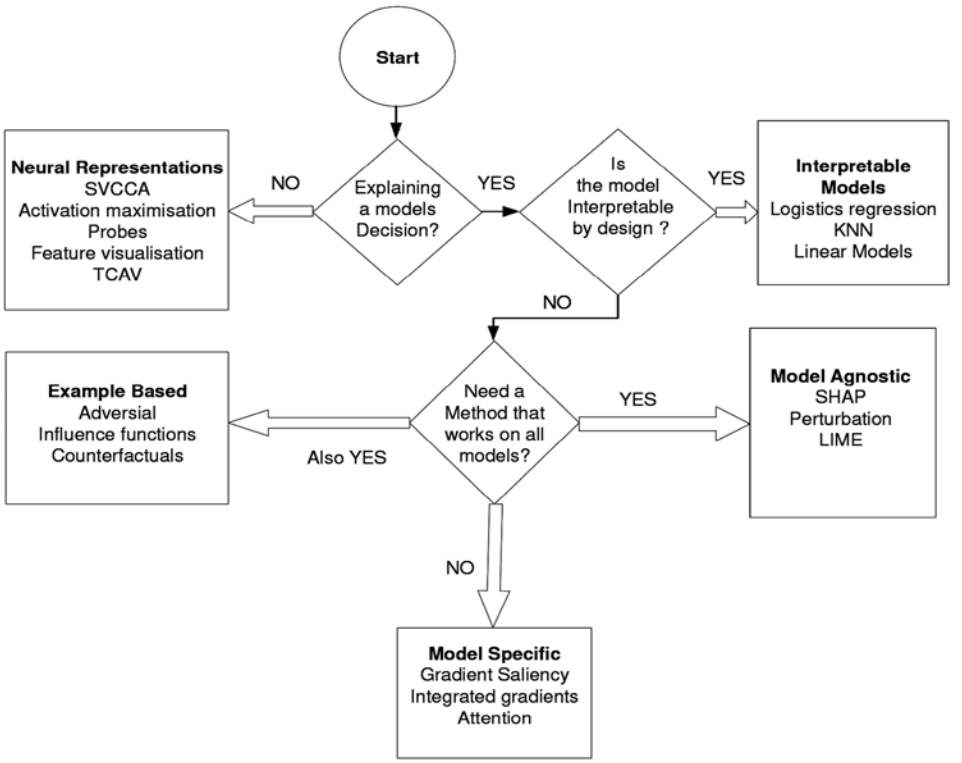
AIX360 or AI Explainability 360 is an extensible open source toolkit that can help you comprehend how machine learning models predict labels by various means throughout the AI application life cycle, and is developed by IBM research.

Skater is a unified framework to enable model interpretation for all forms of models to help one build an interpretable machine learning system, often needed for real world use cases. It is an open source python library designed to demystify the learned structures of a black box model both globally (inference on the basis of a complete data set) and locally (inference about an individual prediction).

For practitioners to successfully select a useful XAI tool, they must be able to assess the tool's ability to provide domain-appropriate explainability, understand its limitations, and have a clear idea about how it can be used to produce meaningful information about ML model predictions.

A high-level guide to the set of tools and methods that helps humans understand AI/ML models and their predictions is referenced in the cheat sheet figure below.

Fig 1: Explainable AI Cheat Sheet[10]



Traceability is a key component of the transparent AI systems. A comprehensive approach to traceability would require, on the one hand, a repeatable execution of the computational steps, but also to capture aspects as metadata that may not be explicit or evident in the digital artifacts.[11]

AI ETHICAL USE

The Lethal Autonomous Weapons Systems (LAWS) technologies are being integrated with AI and ML upfront or at the backend computing systems.

10. Explainable AI Cheat Sheet, at https://jalammar.github.io/explainable-ai/. Accessed on January 3, 2023.
11. Marçal Mora-Cantallops, Salvador Sánchez-Alonso, Elena García-Barriocanal and Miguel-Angel Sicilia, *Traceability for Trustworthy AI: A Review of Models and Tools*, Computer Science Department, Universidad de Alcalá, 28801 Madrid, Spain, at https://www.mdpi.com/2504-2289/5/2/20. Accessed on January 3, 2023.

These are the mainstay aerial weapons whose numbers and lethality will only improve with time.[12] These systems deployed in battle need to be compatible with the operators and must have small error margins.[13] The XAI makes this possible for the users of this technology to understand, appropriately trust, and effectively manage AI.[14]

The North Atlantic Treaty Organisation's (NATO's) six principles of responsible use of AI (military AI should be lawful; responsible and accountable; explainable and traceable; reliable; governable; and have bias mitigation) were adopted in 2020 but with a plan to verify that the principles are followed. Towards which test centres co-located with universities across the various NATO countries would generate knowledge, protocols, tools, etc. to verify this AI deployment.[15]

The Defence Advanced Research Projects Agency (DARPA) in 2016 began to develop 'white box' XAI against the black box problems of traditional AI. The programme was to develop explainability and understandability of AI systems and improve human-machine collaborations.[16] The XAI programme will result in the design of a set of machine learning techniques that create more explainable models while maintaining the prediction accuracy. This is to ensure that the human in the loop understands, appropriately trusts, and effectively manages the developing group of artificially intelligent agents. These agents will have the ability to explain their rationale, characterise their strengths and weaknesses, and convey an understanding of how they

12. Forrest E. Morgan et al, *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World* (RAND Corporation, 2020), at https://www.rand.org/pubs/research_reports/RR3139-1.html#:~:text=A%20steady%20increase%20in%20the,mitigate%20the%20most%2Dextreme%20risks. Accessed on January 3, 2023.
13. Jessica Newman, "Explainability Won't Save AI," The Brookings Institution, May 19, 2021, at https://www.theverge.com/2021/11/8/22769933/israeli-army-facial-recognition-palestinians-track. Accessed on January 4, 2023.
14. Dr. Matt Turek, "Explainable Artificial Intelligence (XAI)", at https://www.darpa.mil/program/explainable-artificial-intelligence accessed on January 4, 2023.
15. Amanda Miller, "NATO's Plan to Grow Trust in Military AI", *Air & Space Forces Magazine*, December 13, 2021, at https://www.airandspaceforces.com/natos-plan-to-grow-trust-in-military-ai/. Accessed on January 4, 2023.
16. Shimona Mohan, "Managing Expectations: Explainable A.I. and its Military Implications", ORF Issue Briefs and Special Reports, August 24, 2022, at https://www.orfonline.org/research/explainable-a-i-and-its-military-implications/#_edn16. Accessed on January 4, 2023.

will behave in the future. These models are envisaged to be combined with state-of-the-art human-computer interface techniques to ultimately translate AI models into understandable and useful explanation dialogues for the end user.[17]

With the enhanced autonomy of aerial weapons and systems, ethical considerations are of great importance. In AI based human decision supported systems or fully automated AI systems, how can responsibility or a chain of accountability for violations be determined, and how can the principles of distinction, proportionality, and humanity be ensured?[18] The Geneva Academy of International Humanitarian Law and Human Rights recommended in 2017 that due exercise of control was necessary on time, to determine and explain the reasoning underlying algorithmic decisions in every combat operation.[19]

The AI ethical principles have been expounded for the development and responsible use of AI and to minimise biases.[20] The reliability and traceability have to be ensured across the entire life cycle of AI capable weapon systems. The control and governance of such systems must retain the ability to disengage or deactivate deployed systems that reveal unintended performance.[21] The war-fighter's trust is to be gained by an education and training, testing, evaluation, verification and validation framework that amalgamates real-time monitoring, algorithm confidence metrics, and user feedback to ensure trusted and trustworthy Al capabilities.[22] AI applications need to be unambiguous well-defined use cases. These must be accurately

17. Ibid.
18. Human Security Centre, "Humans Must Retain Control of Military AI Systems", Posted by Ataa Dabour, March 8, 2021, at http://www.hscentre.org/technology/humans-must-retain-control-of-military-ai-systems/. Acessed on January 4, 2023.
19. Ibid.
20. Deputy Secretary of Defense, 1010 Defense Pentagon, Washington, DC 20301–1010, May 26, 2021, Memorandum For Senior Pentagon Leadership, Commanders of The Combatant Commands, Defense Agency and Dod Field Activity Directors: Subject: Implementing Responsible Artificial Intelligence in the Department of Defense, at https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF. Accessed on January 4, 2023.
21. Ibid.
22. Ibid.

understandable and transparent through the use of review procedures, established with national certification protocols.[23] This is the ultimate war-fighter's XAI.

AI TRUST AND RELIABILITY IN AVIATION

The Boeing 737 MAX crash of October 2018 in Ethiopia was due to the malfunction of the Manoeuvring Characteristics Augmentation System (MCAS), flight control software based on the AI system. MCAS was initially designed as a limited AI system intended to accomplish pitch corrections based on wind resistance and force sensors. However, in 2012, Boeing expanded the MCAS functionality and those changes, coupled with inadequate pilot training, made it more susceptible to catastrophic malfunctioning.[24] This highlighted the airworthiness challenges of AI-based autonomous systems integration into the aerospace domain, particularly the human-machine team dynamics.

During the early 2000 period, the National Aeronautics and Space Administration's (NASA's) Self-Healing Intelligent Flight Control System (IFCS) was designed to adapt to the damage on the aerostructure, stabilise the damaged air vehicle and restore its air handling qualities using adaptive AI technology. It had two AI channels: one, a pre-trained static neural network for the basic aircraft aerodynamic derivatives for each flight condition; and two, a dynamic cell structure neural network to be trained online giving the expected aerodynamic derivatives of the aircraft. The IFCS was, therefore, able to recognise deviations from the reference aircraft behaviour and remodel the flight control laws to recover the reference aircraft behaviour with level one

23. Zoe Stanley-Lockman, Edward Hunter Christie, "Nato Review, An Artificial Intelligence Strategy for NATO", October 25, 2021, at https://www.nato.int/docu/review/articles/2021/10/25/an-artificial-intelligence-strategy-for-nato/index.html. Accessed on January 4, 2023.
24. Alex Henderson Dayton Engineering Advanced Projects Lab, Southwest Research Institute, Beavercreek, USA, Steven Harbour Dayton Engineering Advanced Projects Lab, Southwest Research Institute, Beavercreek, USA, Kelly Cohen, Department of Aerospace Engineering, University of Cincinnati, Cincinnati, USA, "Toward Airworthiness Certification for Artificial Intelligence (AI) in Aerospace Systems". Published in 2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC) September 18-22 2022, at https://ieeexplore.ieee.org/document/9925740. Accessed on January 13, 2023.

handling qualities. The Intelligent Flight Control System Advanced Concept Programme (IFCS ACP), which flowed out of the flight testing of IFCS was a collaboration between NASA and Boeing Phantom Works which led to more research programmes on advanced adaptive flight control systems. A few processes so developed like the state-of-the-art dynamic inversion control technique have been applied in non-adaptive versions on the modern fifth generation fighter aircraft, the Lockheed Martin F-35.[25]

Now, the aerospace industry and the airworthiness authorities are collaborating in introducing the adaptive flight control systems' technologies to the airliners which will provide comfort in the passenger cabins, increased safety, and more robustness to failures and damage.

AEROSPACE VERIFICATION AND VALIDATION

The objective of modern data-centric flight testing is to expand aircraft safety and robust operation in quick time and reduce the cost of testing programmes. Machine learning algorithms are being increasingly deployed to assist flight testers in the extraction and visualisation of patterns that may be obscure to human analysis and anomaly detection.[26] Machine learning embodies applied optimisation algorithms to build models from data. These models are only as good as the data used to train them, and great care must be taken to understand how and when these models are valid. As non-deterministic systems, read AI-based systems, are being introduced into aviation systems, the basic regulatory philosophy will not support the current methods of compliance. In place of a deterministic answer, statistical approaches, and models with dissimilar architectures, or other

25. Rodney Rodriguez, "Geek Culture, NASA's Intelligent Flight Control System, How A.I. Helped NASA to Improve Modern Flight Control Systems", April 18, 2021, at https://medium.com/geekculture/nasas-intelligent-flight-control-system-5dac0a3d3837. Accessed on January 13, 2023.
26. Steven L. Brunton, J. Nathan Kutz, Krithika Manohar, Aleksandr Y. Aravkin, Kristi Morgansen, Jennifer Klemisch, Nicholas Goebel, James Buttrick, "Data-Driven Aerospace Engineering: Reframing the Industry with Machine Learning", *Aerospace Research Central*, vol. 59, issue 8. Published Online: July 20, 2021, at https://doi.org/10.2514/1.J060131, at https://arc.aiaa.org/doi/10.2514/1.J060131. Accessed on January 13, 2023.

The AI systems are also vulnerable to a new type of cyber security attack called an "artificial intelligence attack." Adversaries can sway them to arrive at a wrong conclusion, deceive or shut down decision support to revert to manual modes interfering in the Observe, Orient, Decide, Act (OODA) cycle.

approaches will be required.[27] Flight tests will likely continue to require discrete test conditions. However, an increased emphasis should be placed on analysing data beyond what is required for the test condition and make valid deductions thereof for flight testing objectives. How the AI systems have behaved and what features are responsible for the tester to arrive at an objective conclusion become paramount. The understanding of explainability and traceability cannot be, therefore, lost sight of in such a non-deterministic systems flight-testing regime.

ARTIFICIAL INTELLIGENCE FOR IT OPERATIONS (AIOPS)

The AI systems are also vulnerable to a new type of cyber security attack called an "artificial intelligence attack." Adversaries can sway them to arrive at a wrong conclusion, deceive or shut down decision support to revert to manual modes interfering in the Observe, Orient, Decide, Act (OODA) cycle. General George Patton during the Normandy landing, in World War II, used inflatable balloons painted to look like tanks at *Pas de Calais* to fool the German reconnaissance *Luftwaffe* planes and convince the German command that the invasion point would be *Pas de Calais* rather than Normandy. The German reconnaissance worked by recognising patterns, the shapes and markings representing tanks and other military assets in images, referred to as pattern matching and, hence, was deceived.[28] The

27. Ibid.
28. Marcus Comiter, "Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It," Belfer Centre for Science and International Affairs, Harvard Kennedy School, August 2019, at https://www.belfercenter.org/publication/AttackingAI. Accessed on January 13, 2023.

same shortcoming dooms AI algorithms, to be fooled in pattern recognition neural networks.[29]

Artificial Intelligence for IT Operations (AIOps) is a new domain of AI wherein activities associated with the Information Technology (IT) operations are being managed to ensure uptime, reduce outages and realise service assurance. The developments in the IT infrastructure have moved into software defined resources which require dynamic interventions to reconfigure them on the fly.[30] This is the norm in the network-centric operations so critical for air operations. A vast amount of IT data is now being processed and analysed using AI and ML algorithms to achieve an optimum level of service. This AIOps is as of now narrowly defined and will soon assume a greater role to bring in activities associated with counter-AI attacks either as a subset of cyber warfare or independent operations.

To be a true force multiplier, AI must not only gain trust but also maintain it. In the decision-making OODA loop, where and when to insert human operators to interact with AI systems becomes paramount.

Explainable AI (XAI) is a hinge on which trustworthy AI and AIOps pivot. To be a true force multiplier, AI must not only gain trust but also maintain it. In the decision-making OODA loop, where and when to insert human operators to interact with AI systems becomes paramount. Human operators as system owners are now considered as AI-augmented entities who will continue to be responsible and accountable for its successes and failures.[31] The AI based systems, autonomous or with man in the loop will continue to evolve around the area of trust and its association to transparency and explainability. This is crucial to gaining confidence and, thus, facilitating AI adoption in all aspects of AI enabled systems for air operations.

29. Ibid.
30. Moogsoft Platform, "An Everything Guide to AIOps", at https://www.moogsoft.com/resources/aiops/guide/everything-aiops/. Accessed on January 13, 2023.
31. Bob Friday, Juniper Blogs, "Explaining AI is Key to Trusting AI and Adopting AIOps", September 28, 2022, at https://blogs.juniper.net/en-us/industry-solutions-and-trends/explaining-ai-is-key-to-trusting-ai-and-adopting-aiops. Accessed on January 13, 2023.

AI TRUST-BASED COMMAND AND CONTROL (C2)

Use of AI allows appreciating the operating air environment in greater detail and, hence, reduces the cognitive load on the commander's decision-making. This is an evolution, and AI-enabled decision-making now calls for a new approach in command and control structures. "Whatever our C2 models, systems and behaviors of the future will look like, they must not be linear, deterministic and static. They must be agile, autonomously self-adaptive and self-regulating, and at least as contingent and emergent as the context within which they are formed and operate."[32]

Commanders are not expected to be experts of AI but need to be proficient to interact with technology. They are to remain focussed on the operational requirements that AI intends to serve and be satisfied enough to trust the AI outputs for decision-making. The commander must retain his inquisitive attitude in the process and not be merely a recipient of algorithm output. Future air commanders will lead diverse inter-disciplinary teams to understand the complex air environment. AI agents being dovetailed into the team changes the team dynamics of the human-machine team formation.[33] How the concept of AI as a team member will unfold remains unclear. Explainability will ensure that both remain on a more equal footing with commanders as leaders and warriors, not merely people who are good managers of the battle. XAI is a tool to maximise the chances of the armed forces winning in a deeply competitive environment. The trust on such algorithms should never be absolute but must be referenced to different elements called trust points of deployment, data, process, output, and organisational system trust. This is the overall ecosystem for optimising the use of AI. The quality of data or the explainability of AI outputs should not become a misplaced sense of reassurance about the technology.[34]

32. Multinational Capability Development Campaign, "Final Study Report on Information Age, Command and Control Concepts", February 8, 2019, p. 20, at https://www.innovationhub-act.org/sites/default/files/AxS_Product_2light.pdf. Accessed on January 14, 2023.
33. Christina Balis and Paul O'Neill, "Trust in AI: Rethinking Future Command", June 23, 2022, at https://www.qinetiq.com/-/media/59de8c79f10d4c5699b5a1f25f2d2279.ashx. Accessed on January 14, 2023.
34. Ibid.

MILITARY AI EDUCATION STRATEGY

The people who are associated in the use of AI technologies are critical in the optimum employment of AI capabilities. AI education strategy as a high-level learning journey is, therefore, important to ensure that the military can win future engagements with the adversaries. The workforce training will have to cater to both technical and non-technical skill sets across all positions and ranks in integrating organisational AI war-fighting capabilities. Leadership support is required for a multi-pronged strategy that involves end users (employ AI) as well as middle managers (drive AI).[35] At this level, individuals use AI-enabled systems to identify threats on the battlefield and need to understand how data collection and curation affect the outcomes and system accuracy. Therefore, training at this level must remain focussed on the unique needs and potential contributions of each archetype. On the other extreme are create AI and embed AI archetypes which require advanced levels of training, including an advanced degree.[36]Attaining these qualifications is intensive to acquire the level of desired mastery.

An AI trade institute within the service would be able to accomplish the required education and training for many of these applied competencies tailored to meet the trade requirements. It can also be used to teach the foundational and applied topics that provide leaders with the tools they need to understand AI capabilities which can be integrated into the mission ethically and responsibly. Similarly, courses and workshops focussing on the deployment and responsible application of AI, AI capabilities impacting organisation, oversight of AI-enabled systems, leading practices for human-machine interaction, engaging with, and interpreting, AI applications, etc.[37] The canvas is wide and dynamic to indoctrinate the air warriors in employment of AI enabled systems. The Indian Institute of Technology (IIT)

35. Robert Beveridge, "How to Grow an AI-Ready DoD Workforce", September 12, 2022, at https://insights.sei.cmu.edu/blog/how-to-grow-an-ai-ready-dod-workforce/. Accessed on January 14, 2023.
36. Ibid.
37. Ibid.

in Mandi has signed a pact with the Indian Air Force's (IAF's) Headquarters Maintenance Command, Nagpur, to collaborate on research projects, technology development and skill development in the areas of artificial intelligence, machine learning, human-computer interaction, and decision support systems.[38]

A large pool of available talent in India has encouraged many world-renowned AI research institutes for establishing AI labs and research centres. The various IITs, Defence Research and Development Organisation (DRDO) and private industry in collaboration with IIT/National Institute of Technology (NIT)/Indian Institutes of Science (IISc) have established AI innovation hubs/research facilities in various AI domains. The qualified youth in India are joining these reputable AI research centres to contribute for the benefit of the country.[39] The Indian Air Force Centre of Excellence for Artificial Intelligence under the aegis of UDAAN (Unit for Digitisation, Automation, Artificial Intelligence and Application Networking) has been established at Air Force Station, Rajokri, New Delhi, since July 9, 2022, and the Al based applications are being developed with in-house expertise in coordination with various Public Sector Undertakings (PSUs), Medium, Small and Micro Enterprises (MSMEs) and leading academia.[40]

RECENT ADVANCES

As has been discussed in this article so far, XAI provides explanations of black box models, to reveal the features of the AI model that have resulted in the decision-making using various tools, techniques, and algorithms. Human understanding is vastly facilitated through various visualisation

38. "IIT Mandi Inks Pact with IAF's HQ Maintenance Command for R&D in AI, Machine Learning", January 23, 2023, at https://www.tribuneindia.com/news/himachal/iit-mandi-inks-pact-with-iafs-hq-maintenance-command-for-rd-in-ai-machine-learning-473010. Accessed on January 14, 2023.
39. Nivash Jeevanandam, India AI, "Interesting AI Research Labs in India", at https://indiaai.gov.in/article/interesting-ai-research-labs-in-india. Accessed on January 14, 2023.
40. Manish Kumar Jha, "IAF Centre of Excellence for Artificial Intelligence to Redefine Air Combat", Updated: July 12, 2022, at https://www.financialexpress.com/defence/iaf-centre-of-excellence-for-artificial-intelligence-to-redefine-air-combat/2589934/. Accessed on January 14, 2023.

techniques presented in a more explainable and interpretable way.[41] Visual Analytics (VA) is a field of study to portray the data/model in an interactive visual interface, especially for those uninitiated in ML. VA combined with XAI algorithms will be a great stride in the explainability task for the front-end users of AI agents.[42] Visual-based XAI (vXAI) is an emerging research area in XAI which will be a real game changer in the use of AI in the military aviation space. The IAF can assess the research direction to the AI applications being developed for better interpretation of neural networks and more transparency through XAI methods in the field of visual analytics.

Reinforcement Learning (RL) methods of AI are increasingly becoming a mainstay for the cognitive models in the progress of XAI. RL is being extensively deployed in various AI agents of aerospace autonomous control of air vehicles and gaining maturity to form part of manned-unmanned air team combat and, hence, the explainability and trust of RL in air war decision-making is important. Explainable RL (XRL) is a recent field of research that aims to develop techniques to extract concepts from the AI agents for XAI deployment so that the agents can answer questions that explain outcomes and justify decisions.[43] As of now, most XRL methods are imitating and simplifying a complex model instead of designing an intrinsically simple one. Thus, inter-disciplinary research work is necessary to alter the generated XAI explanations for a non-AI expert human user.[44]

Air Traffic Management (ATM) is now an increasingly complex arena of massive air traffic density of civil airlines, with military aviation being asked to share space in pre-defined pockets. The plethora of manned and unmanned air vehicles rapidly contesting the congested complex air space

41. Gulsum Alicioglua, "A Survey of Visual Analytics for Explainable Artificial Intelligence Methods", *Computers & Graphics*, vol 102, February 2022, at https://www.sciencedirect.com/science/article/abs/pii/S0097849321001886. Accessed on January 21, 2023.
42. Ibid.
43. Richard Dazeley, Peter Vamplew, Francisco Cruz, "Explainable Reinforcement Learning for Broad-XAI: A Conceptual Framework and Survey", at https://arxiv.org/abs/2108.09003. Accessed on January 21, 2023.
44. Erika Puiutta and Eric MSP Veith, "Explainable Reinforcement Learning: A Survey", at https://arxiv.org/abs/2005.06247. Accessed on January 21, 2023.

Air Traffic Management (ATM) is now an increasingly complex arena of massive air traffic density of civil airlines, with military aviation being asked to share space in pre-defined pockets.

promises AI a great potential for enabling sustainability. However, in this safety critical industry, decision-making AI needs "explainability" which is an essential and critical requirement in improving trust and reliability in the interaction between human Air Traffic Control (ATC) controllers and AI.[45] The ARTIMATION project by the Artificial Intelligence/Computer Science at Mälardalen University, Sweden, is using AI sophisticated techniques to assist controllers to visualise the decision-making processes in the OODA loop. It has two tools: one, 3D conflict detection; and two, a resolution tool that has been developed with XAI algorithms to understand and accept the XAI outcomes.[46]

AI can definitely boost ATM performance by increasing capacity, reducing delays and improving safety but since controllers are eventually responsible for their decisions, building trust remains fundamental to AI adoption. A more all-embracing TAPAS (Towards an Automated and explainable ATM System) project, funded within the SESAR Joint Undertaking, a public-private partnership has been set up to modernise Europe's ATM system.[47] It has XAI prototypes combined with visual analytics, to obtain information from the AI agents and display it on-screen in user-friendly formats. XAI is being used in both non-safety-critical air traffic flow and capacity management and safety-critical conflict detection and resolution. TAPAS employs XAI techniques which explain decisions taken by the AI system.[48]

45. SESAR Joint Undertaking, "Explainable Artificial Intelligence—Why it Matters for Air Traffic Management", March 21, 2022 at https://www.sesarju.eu/news/explainable-artificial-intelligence-why-it-matters-air-traffic-management. Accessed on January 21, 2023.
46. Ibid.
47. Cordis, "Transport and Mobility, Explainable AI for Increased Trust in Air Traffic Management Software", October 7, 2022 at https://cordis.europa.eu/article/id/442209-explainable-ai-for-increased-trust-in-air-traffic-management-software. Accessed on January 21, 2023.
48. Ibid.

AI in civilian ATM is still in its infancy and most airport Primary Surveillance Radars (PSRs) used for air traffic control, do not perform real-time classification of aircraft radar signatures. This is an exciting opportunity in the field of XAI in radar systems to develop deep learning classification and explainability methods that can be applied to a safety-critical radar system for ATM.[49] This automated way of mapping radar signatures to discrete aircraft classes can also be migrated to airborne weapon radars albeit with different operational imperatives. The IAF may consider funding such research projects in XAI solutions, ranging from data feature-based to symbolic-based to explain the deep learning algorithm actions. Despite copious research on technology, the practical implementation of XAI is lacking due to various regulatory requirements, practical challenges in its implementation and transparency.[50] These need to be harmonised and given the impetus by the users of AI applications and decision support systems.

The IAF may consider funding such research projects in XAI solutions, ranging from data feature-based to symbolic-based to explain the deep learning algorithm actions.

In conclusion, it is generally agreed that the future of air combat is trustworthy and scalable autonomous AI systems. The conduct of air operations will be dividing the air elements into smaller, less expensive units, with fewer people, more expendable machines—"mosaic warfare" as elucidated by DARPA.[51] The human operators must trust the AI system in order to be able to perform their future role as "battle managers". Any degradation of trust will call for a human takeover.[52] The man-machine

49. "Explainable AI for Radar in an Air Traffic Management System", at https://www.findaphd.com/phds/project/explainable-ai-for-radar-in-an-air-traffic-management-system-phd/?p125124. Accessed on January 22, 2023.
50. 2021 IEEE/AIAA 40th Conference on Digital Avionics Systems (DASC), "An Explainable Artificial Intelligence (xAI) Framework for Improving Trust in Automated ATM Tools", at https://ieeexplore.ieee.org/document/9594341. Accessed on January 22, 2023.
51. The Decoder AI Research, "Artificial Intelligence to Fly US Fighter Jets in 2024", February 12, 2022, at https://the-decoder.com/artificial-intelligence-to-fly-us-fighter-jets-in-2024/. Accessed on January 23, 2023.
52. Ibid.

interface is to be a harmonious teaming effort of AI agents with explainability of actions a mandatory necessity for the team. Visual and other interfaces will be interactive with the AI agent keeping the human in the loop informed of the AI-based decisions enhancing his trust. The Explainable AI (XAI) domain, therefore, needs to be fully imbibed in the development and employment of AI-based air systems.