ADVERSARIAL AI CONCERNS IN AIR WARFARE PLATFORMS

SUKHCHAIN SINGH

The sad thing about artificial intelligence is that it lacks artifice and, therefore, intelligence.

– Jean Baudrillard

Artificial Intelligence (AI) is a disruptive technology for military commanders to reduce their own kill chain time and gain advantage in the battlefield Observe-Orient-Decide-Act (OODA) cycle. However, the embedded AI in systems and modules using the deep learning algorithms are susceptibile to the risk of being interfered in by the enemy and, hence, have to be protected. Similarly, the opportunity of delaying or obstructing the adversary's kill chain of AI aided weapon systems as a military strategy must not be fettered away. These adversarial actions are deceptions for the AI models at various steps and activities of the weapons kill chain resulting in unreliability, and generating mistrust on the commander's ability to effectively use the technology for the intended objective.[1]

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 and M Tech from the Indian Institute of Technology (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).

Adversarial examples in computer vision algorithms include injection of minuscule noise that is difficult to distinguish by humans, but can cause AI algorithms to generate wrong decisions. Specifically, in the field of image recognition AI algorithms, the attack of adversarial samples will adversely affect the accuracy of the model. Adversarial examples in natural language processing, audio recognition, deep learning and reinforcement learning algorithms used in military air platforms have been reported by researchers in the laboratory as well as real physical systems, including autonomous driving, face recognition, object detection and robot navigation.[2] The security of own AI in weapon systems combined with how to exploit the AI adversarial actions on the enemy is an important research focus area in the deployment of AI in all weapon systems, particularly in air warfare platforms.[3]

Conversely, RAND has propounded that adversarial attacks intended to deceive, require an intimate knowledge of the system and are operationally infeasible to design and deploy. Thus, such actions pose less risk than what academic researchers have hyped. Nonetheless, well-crafted AI systems, with adversarial mitigation strategies, can further reduce the risks of such attacks. However, how the model can be influenced by the adversary needs to be understood and knowledge leaks of the AI model plugged at various stages of its development and deployment. One cannot be complacent about the developing technology and must remain abreast of the academic state-of-the-art methods to attack AI in real-world scenarios and understand how

1. Yuwei Chen, "The Risk and Opportunity of Adversarial Example in Military Field", Chinese Aeronautical Establishment. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2022, pp. 100-107, https://openaccess.thecvf.com/content/CVPR2022W/ArtOfRobust/papers/Chen_The_Risk_and_Opportunity_of_Adversarial_Example_in_Military_Field_CVPRW_2022_paper.pdf
2. Richard Tomsett, Amy Widdicombe, Tianwei Xing, Supriyo Chakraborty, Simon Julier, Prudhvi Gurram, Raghuveer Rao and Mani Srivastava. "Why the Failure? How Adversarial Examples can Provide Insights for Interpretable Machine Learning", 21st International Conference on Information Fusion (FUSION), 2018, pp. 838–845, 2018, https://discovery.ucl.ac.uk/id/eprint/10070702/1/2374_paper.pdf
3. Chen, n. 1.

these technologies practically affect our own concepts of operation and our adversaries.[4]

AI-ASSISTED AIR POWER

AI augmented system developments have been evolving in air power and have a profound impact on war-fighting efficacy. Important disruptive applications of AI are the autonomous fighter aircraft, drone-wingmen, swarms, decoys, and the innovative avatar aviators.[5]

The autonomous platforms' AI algorithms, as demonstrated by the Defence Advanced Research Projects Agency (DARPA) (2023) and its noteworthy Air Combat Evolution (ACE) programme have, in just three years, transited from computer simulated F-16 aerial dogfights to real-time physical controlled dogfights in a controlled air environment. AI agents can now safely manoeuvre an actual fighter in flight. Like the F-16 AI agent, the avatar aviator has to asses when to take over the operational control of the aircraft from the overloaded human pilot and allow him to concentrate on the core combat related tasks. And, in an extreme case of incapacitation of the pilot, autonomously recover the aircraft safely.[6]

The Tempest is an AI-driven innovation of 'loyal wingmen' which is an autonomous Unmanned Aerial System (UAS) flying alongside a manned fighter or as a decoy to protect the main fighter from enemy air defence weapons. Similarly, the MQ-28A Ghost Bat is the Boeing's Airborne Teaming System which is an AI-assisted semi-autonomous stealthy loyal wingman. These AI-assisted teaming systems will ensure enhanced lethality and

4. Li Ang Zhang, Gavin S. Hartnett, Jair Aguirre, Andrew J. Lohn, Inez Khan, Marissa Herron, Caolionn O'Connell, "Operational Feasibility of Adversarial Attacks Against Artificial Intelligence", RAND NATIONAL SECURITY RESEARCH DIVISION. December 12, 2022, https://www.rand.org/content/dam/rand/pubs/research_reports/RRA800/RRA866-1/RAND_RRA866-1.pdf
5. Professor Ron Matthews, "Disruptive AI, The Accelerating Impact on Next-Generation Air Power", 11th DIACC (Dubai International Air Chiefs Conference), November 12, 2023, *Air Power Journal*, 2023, selected papers, https://www.diacc.ae/wp-content/uploads/2023/10/P1-Professor-Ron-Matthews-Disruptive-AI-Tawazun-www.diacc_.ae-www.theairpowerjournal.com-www.spps_.se_.pdf
6. Ibid.

Air warfare engages, and is moulded by, technology. The operational data is gleaned from a plethora of air platforms, and routed, stored and processed in remote or edge AI processors.

survivability in contested air environments for the manned fighters and next generation air dominance platforms.[7]

DARPA's "Autonomous Multi-Domain Adaptive Swarms-of-Swarms" (AMASS), as per the Pentagon programme, comprises the utilisation of swarms of swarms to develop a counter-Anti-Access/Area Denial (A2/AD) capability. These are again AI-enabled autonomous drone swarm systems controlling other swarms which could overpower enemy defences.[8]

The Miniature Air-Launched Decoys (MALDs) are AI driven autonomous decoys with an exceptional ability to impersonate US or allied aircraft in the air and deceive the most advanced air defence systems. These have been designed to "deceive, distract and saturate" radar systems with deceptive signals mimicking the F117 stealth fighter or B-52 bombers. When deployed in an area with numerous aircraft and other airborne platforms, the adversary air defence systems are forced to differentiate between real and fictional radar returns for their tactical actions.[9]

Air warfare engages, and is moulded by, technology. The operational data is gleaned from a plethora of air platforms, and routed, stored and processed in remote or edge AI processors. The proliferation of mobile systems, Unmanned Aerial Vehicles (UAVs) and Remotely Piloted Vehicles (RPAs) for remote sensing are part of the Internet of Battlefield Things (IoBT) that collaborate in swarms to generate the air situation and environment. This is the enabling technology for emerging operational concepts of 'mosaic' for multi-domain operations. It is no longer a linear kill chain fixed data flow, rather a fluid and varying data flow in the IoBT communication architecture.[10]

7. Ibid.
8. Ibid.
9. Ibid.
10. Dr Peter Layton, "Future Options for Artificial Intelligence and Machine Learning Assisted Decision-Making in Air Warfare", *Air Power Journal*, First Edition (2021), Chapter 4 in Multi-

Thus, to conduct a multi-domain operation, it is imperative to employ AI-assisted automated systems for quick decisions involving complex scenarios and identify, localise and recognise objects across the battlespace.

AI assisted systems are excellent at discovering hidden patterns within a high clutter background, however, as discussed earlier, they can be fooled due to lack of robustness.

AI is not flawless in its process and can be deceived due to its inherent learning mechanism and, hence, cannot transfer the obtained knowledge between various tasks. The future air battlespace will be cluttered with numerous air platforms tasked with varying air operational roles as well as AI-based Electronic Warfare (EW) systems to deliberately confuse and mislead the enemy in the conflict arena. AI assisted systems are excellent at discovering hidden patterns within a high clutter background, however, as discussed earlier, they can be fooled due to lack of robustness.[11] The air commander must be cognisant of this aspect of AI in deployment and tailor his plan of responses appropriately.

AI THREAT ANALYSIS

What is the threat of military AI systems being interfered with and deceived? Let us visualise the following decisions made by an AI autonomous weapon platform under a mission wherein military targets of interest like combat vehicles, enemy fighter aircraft and a platoon of soldiers are recognised as a hut, a bird or a herd of animals; or own convoy misidentified as adversary armoured vehicles on the move; or a benign gathering of people in a playground photographed through reconnaissance systems is misinterpreted as armoured personnel carriers. Decisions made through such compromised systems autonomously or with the man in the loop can be catastrophic mission objective failures and embarrassing for the

Domain Operations, Artificial Intelligence and Information Dominance, https://www.diacc.ae/resources/2021_Peter_Layton_AI_ML_Assisted_Decision_Making_Air_Warfare.pdf

11. Ibid.

commander. The manned-unmanned teaming of fighter aircraft, remotely piloted vehicles, smart autonomous ammunitions with seek and destroy capabilities, airborne autonomous reconnaissance and surveillance systems, integrated air command and control systems for Air Defence (AD) all of which now incorporate AI capabilities, if deceived through adversarial AI, can erode the trust of the commander, forcing him to revert to manual intervention, thereby losing the advantage offered by the AI technology for his OODA loop. This is the evolving field of adversarial AI in the academic and military arenas. A plethora of adversarial AI examples are reported in various research papers and discussed for their utility in the military domain as well.[12]Automated reconnaissance camera systems can be disabled, enabling the intruders to enter without being detected; the navigation systems within autonomous air vehicles can be compromised, leading to unintended manoeuvres, flight routes being modified away from waypoints or targets, or causing a controlled flight into the terrain.[13] AI-assisted systems undertake tasks that classically involve human intelligence. They utilise a variety of technologies, including Machine Learning (ML) models, high performance computer hardware, a plethora of different sensors, software, robotics and other statistical algorithms. ML is a subset of AI that is a statistical model, which learns from training data.[14]

A hypothetical real-time mission has been enumerated in Jacob Simpson's paper[15]which brings out the strength of the Intelligence, Surveillance and Reconnaissance (ISR) artificial intelligence system. The air commander in an Integrated Air Command and Control System (IACCS) is monitoring the unmanned strike force being pushed into the enemy territory to destroy a

12. Dr Elie Alhajjar, "Adversarial Machine Learning Poses a New Threat to National Security", Armed Forces Communications & Electronics Association International (AFCEA), Cyber Edge, July 1, 2022, https://www.afcea.org/signal-media/cyber-edge/adversarial-machine-learning-poses-new-threat-national-security
13. Richard Flint, "Military Adoption of AI is Threatened by a New, Intelligent Threat", Linkedin, June 6, 2022, https://www.linkedin.com/pulse/military-adoption-ai-threatened-new-intelligent-threat-richard-flint
14. Ibid.
15. Jacob Simpson, "Operations in Deception: Corrupting the Sensing Grid of the Enemy, *The Forge*, The Australian Joint Professional Military Education (JPME) Continuum Article, https://theforge.defence.gov.au/article/operations-deception-corrupting-sensing-grid-enemy

target of vital importance and surprise is essential for the success of the mission. The mission and the air situation are being closely monitored when an alarm is flashed on the screen to indicate that an unknown Surface-to-Air Missile (SAM) system radar has been switched ON. Has the surprise been lost for the mission, is rerouting of the package required, is it a GO? The air commander initiates a query on the real-time data set of enemy air defences and weapons through the AI system of IACCS. The AI system algorithm generates a new route and the AI assesses no significant enemy posture to give the confidence to continue with the mission in the enemy air defence domination area without any loss of surprise. The example also brings out the mischief which can be played into the sense grid of the command and control system. The AI systems are trained to detect anomalies by utilising training data, and for the sense grid, these are simulated since no actual data exists for the wars to be fought in the future domain. Generative AI, thus, assumes importance to create such data, and should be independent of beliefs, culture and the biases of those that create them.[16] The counter-AI strategy is to create mistrust in the enemy's own AI systems by exploiting his sense grid vulnerabilities. Even if it is done once, it will force him to rethink at all times, thus, slowing down his OODA loop and negating the AI advantage. Similarly, one has to guard against this for own AI systems to retain an advantage over the adversary.[17]

ATTACKING AI

Attacking AI systems is a complex study of computer sciences and attempting to simplify this has its pitfalls in realistic understanding. However, a plethora of academic research and literature is available online for the more tech-savvy reader. Marcus Comiter's paper[18] and the US Department of

16. Ibid.
17. Ibid.
18. Marcus Comiter, "Attacking Artificial Intelligence", Belfer Centre for Science and International Affairs, PAPER, August 2019, https://www.belfercenter.org/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf

Homeland Security, Science and Technology study[19] are such write-ups, which are quite exhaustive in their coverage.

The adversarial machine learning objective is to trick machine learning models by inserting input to deceive the AI system. This includes both the generation and detection of adversarial examples. These examples comprise input data specifically crafted to deceive AI classifiers, e.g. in image recognition, where alterations are done on images that cause a classifier to make inappropriate predictions. These adversarial examples are inputs which resemble a valid input to the learning model that are intentionally devised for a model to make a mistake in its predictions.[20] When the target model's complete architecture and parameters are known, adversarial AI attacks are white box attacks, but when the attacker can only observe the target model's outputs, then the black box attacks are attempted.[21]

Adversarial attacks are classified as poisoning attacks, evasion attacks or model extraction attacks. AI systems can be re-trained using data collected during operations. The attacker may poison the data by injecting malicious samples during operations, which subsequently disrupt or influence re-training. In evasion attacks, the attacker manipulates the data during deployment to deceive previously trained classifiers. Model stealing or model extraction involves an attacker probing a black box machine learning system in order to either reconstruct the model or extract the data it was trained on. This is especially significant when either the training data or the model itself is sensitive and confidential.[22]

Some popular adversarial attack methods which are based on AI algorithms in use are:[23]Limited-memory BFGS (L-BFGS), Fast Gradient Sign Method (FGSM), Jacobian-based Saliency Map Attack (JSMA), Deep

19. "Risks and Mitigation Strategies for Adversarial Artificial Intelligence Threats", A DHS S&T Study, Preparedness Series, June 2023, USA Department of Homeland Security, Science & Technology, https://www.dhs.gov/sites/default/files/2023-12/23_1222_st_risks_mitigation_strategies.pdf
20. Gaudenz Boesch, "What Is Adversarial Machine Learning?" Attack Methods in 2024, Viso.ai blog, https://viso.ai/deep-learning/adversarial-machine-learning/
21. Ibid.
22. Ibid.
23. Ibid.

Fool Attack, Carlini & Wagner Attack (C&W), Generative Adversarial Networks (GAN), Zeroth-Order Optimisation (ZOO) attack. They have their own strengths and weaknesses. New and innovative attack methods are continually being developed.

The methods supporting the artificial intelligence systems are now susceptible to a new form of cyber security attack called an "AI attack." The adversary can influence these systems to alter their behaviour to serve a malicious end goal. The conventional cyber attacks are initiated by inserting "bugs" in the programme code but the AI attacks are facilitated by intrinsic constraints in the resident AI algorithms, a feature of how they learn, and can be attacked and deceived. Now for AI attacks, physical objects can be used in cyber warfare (e.g. altering a stop sign into a green light using a piece of tape for fooling the autonomous car). Training data can also be exploited similarly to deceive and train misappropriately, necessitating a relook at how data is gathered, warehoused, and deployed. These cyber security problems cannot be resolved using the existing policies and tool kits. AI attacks cannot be 'fixed' or 'patched'. A completely new approach to mitigate, address and protect is needed which will require new technical solutions.[24] DARPA and the Air Force Research Laboratory (AFRL) have released the moving and stationary target acquisition and recognition data set[25], which is being deployed in AI techniques to classify and recognise targets of interest.

Air forces operate in a unique environment. The adversary can develop unique ways to attacks the AI-based airborne platforms or ground-based assets. And, hence, there are unique challenges in defending them. The loss and capture of aircraft, drones and weapon systems, on which AI systems will be deployed, is an accepted reality in war. With edge computing, the data and AI algorithms are stored and run directly on these platforms for optimum utilisation of the communication bandwidth, thus, their capture will compromise the AI model and its capability. The AI systems, therefore,

24. Comiter, n. 18.
25. MSTAR Public Targets, https://www.sdms.afrl.af.mil/index.php?collection=mstar&page=targets.

AI attacks will be difficult to detect during operations because the adversary would be only interested in scouting to learning the data sets or types of tools being used which itself may not trigger any alarms unless a serious breach occurs.

have to be accorded the same protection category as physical assets requiring self or command-initiated destruction.[26]

Multiple military AI systems will use shared data sets. These are expensive, require domain expertise, are time-consuming and difficult to create. One instance of compromise in any system of data set would render all other dependent systems vulnerable to attack. Also, if this data is hacked, then every AI application developed using this data would be potentially compromised. Therefore, when developing AI-enabled weapons and defence systems, the individual data samples used to train the models themselves become a secret that must be protected. These best practices are necessary in the military AI field, and require new policies for managing data acquisition and preparation.[27] The models and tools will be targets for adversaries to steal through hacking or counter-intelligence operations. Traditional cyber actions will be utilised to obtain the models stored running on these systems which the adversaries can back-solve for the attack patterns that would deceive the systems.[28]

AI attacks will be difficult to detect during operations because the adversary would be only interested in scouting to learning the data sets or types of tools being used which itself may not trigger any alarms unless a serious breach occurs. The networks are still based on traditional cyber attack responses, something that has now to be revisited. The detection of an intrusion into assets of AI training models must become more robust and if an intrusion is detected, those assets may have to be shut down or retrained. Another approach could be to continually monitor the AI systems' performance to detect an AI attack on the assets. Though AI attacks can undoubtedly be launched without associated cyber attacks, strong traditional

26. Comiter, n. 18.
27. Ibid.
28. Ibid.

cyber defences will ensure increased effort in crafting AI attacks.[29]

An understanding of the AI models' pattern recognition should help the adversary to launch a well-planned attack on the system or even prevent it by own platforms. The deep neural networks are like black boxes whose output explainability and the mechanics of coming to a particular decision are still not fully understood. Therefore, what is not understood cannot be fixed reliably and it becomes difficult to asses if the model is not functioning properly or has been compromised in the AI attack. Algorithms based on decision trees and regression models are fully understandable but AI systems based on such models are not comparable to the high performance demonstrated by neural networks. That is why these characteristics are the reasons why there are no perfect technical fixes for AI attacks.[30]

"AI Suitability Tests" must be conducted to ascertain the present and future risks in deployment of AI in military systems, particularly autonomous airborne platforms.

"AI Suitability Tests" must be conducted to ascertain the present and future risks in deployment of AI in military systems, particularly autonomous airborne platforms. The quantum of weightage given to AI in the functioning of the system application must be gauged based on its attack vulnerability and the ramification on its reliability to perform the intended task. A non-AI system may be considered in its place for that application of importance.[31]

AI appears to be human, resembling Frankenstein's monster, but is unequivocally not human. The AI algorithms are taught pattern matching of the various data they work on and, hence, are inherently susceptible to influence and poisoning during all phases of their practice: from how they learn, what they learn from, and how they operate. The unconstrained deployment of AI into the military and society is fanning future vulnerability. Thus, the military planners must consider addressing AI security compliance

29. Ibid.
30. Ibid.
31. Ibid.

programmes in all weapon systems.[32] For AI, decision-makers have been urged to adopt this technology with greater nuance and cynicism, because the complexity of systems integration and test processes will naturally limit its transformative potential for combat systems.[33]

TEST AND EVALUATION OF AI

The stochastic capabilities of AI which are data-centric, black box approach, self-learning capabilities and with an adaptive nature are a challenge for the Test and Evaluation (T&E) processes for evaluating air platforms and associated on-board systems. The difficulty of AI T&E is further compounded by the on-boarding of hybrid weapon systems which are an amalgamation of legacy non-AI systems, new non-AI systems, contemporary or legacy systems with AI that are 'bolted on', and AI that is 'baked-in', functioning concurrently as a cohesive package in the air platform. The final call is always to test the system's performance as per the laid down baseline capabilities.[34]

For T&E of the adversarial attack, the attack surface is laid out in the National Institute of Standards and Technology (NIST) (NIST SP 800-172 from GAO-19-128) for all systems, including AI-assisted systems. These boundary points include the attacker's possible areas of penetration, the affected locations in the architecture, or data extraction from the system compromised. These are also the same surfaces for the cyber attack and used by the tester, which can be exploited by the adversary to interrupt, deny, or degrade the performance of the AI system.[35]

32. Ibid.
33. Dr Ted Harshberger and Dr Cynthia R. Cook, "The Elusive Promise of Digital Acquisition for Combat Capabilities", 11th DIACC (Dubai International Air Chiefs Conference) November 12, 2023, *Air Power Journal 2023*, selected papers, https://www.diacc.ae/wp-content/uploads/2023/10/P6-Dr.-Ted-Harshberger-Dr.-Cynthia-Cook-The-Elusive-Promise-of-Digital-Acquisition-CSIS-www.diacc_.ae-www.theairpowerjournal.com-www.spps_.se_.pdf
34. "Test and Evaluation Challenges in Artificial Intelligence-Enabled Systems", National Academies of Sciences, Engineering, and Medicine, 2023, Department of the Air Force. Washington DC: The National Academies Press, https://doi.org/10.17226/27092.
35. Ibid.

However, elaborate surface mapping is needed due to the data-centricity of the AI-based algorithms in the systems. Thus, the tester would be required to look into corruption in the entire life-cycle of AI systems, including the training data used, the testing, operations and the access to the details in any configuration or model that is part of the AI component. The adversarial attack in the supply chain of the software and data used to construct the AI component becomes very relevant. Therefore, testing the AI components as a standalone system has to be initiated at the basic level before undertaking the entire AI interactive chain of the platform T&E.[36]

An incident of an alleged AI-assisted drone of the US Air Force (USAF) killing its own operator caused shivers and raised alarm about the safety of AI: was it an adversarial attack? This was later denied, but it reinforces that AI can go rogue, causing unintended consequences.[37]

Measures like adversarial training, platform security updates, reliable auditing, and data cleansing can improve the robustness of AI models against directed attacks. The strategy must be a step ahead of adversaries through research, proactive approach and collaboration with academia in developing AI technology.[38] Apprehensions about the vulnerabilities and risks associated with the retrofit integration of AI into legacy digital technologies infrastructure must be ruthlessly analysed. One has to be cautious about the T&E of the safety of the flight or mission critical systems.[39] The Aircraft Systems Testing Establishment (ASTE) and Software Development Institute (SDI) of the Indian Air Force (IAF) are involved in the T&E and development of software of air platforms being inducted into the air force. ASTE must now seriously consider having an AI T&E domain specialist group in its fraternity of test pilots and flight test engineers. SDI software specialists may be coopted in the testing programme of the AI-based systems under flight evaluation. AI will permeate all airborne systems, therefore, detailed

36. Ibid.
37. "USAF Official Says He 'Misspoke' About AI Drone Killing Human Operator in Simulated Test", June 1, 2023, ADVERSA, AI Blogs, https://adversa.ai/blog/towards-trusted-ai-week-23-ai-drone-killing-and-adversarial-attacks-in-military/
38. Ibid.
39. Ibid.

and updated AI domain foundational knowledge should be mandatory in the training curriculum of the Air Force Test Pilots Flight School (AFTPS). Exchange programmes and specialist knowledge courses with the US Air Force Test Pilot School (USAF TPS)and the flight-testing establishments may be considered to generate the requisite flight-testing schedules at this stage of AI technology proliferation into air platforms.

RISK MITIGATION

By engaging security researchers to inspect AI models, similar to ethical hackers testing cyber defences, the cooperative knowledge can reinforce AI's security. This may be the desirable approach to prevent adversarial attacks later in the deployment phase.[40]

In India, this is the time to address the rise of adversarial AI. Public Private Partnerships (PPPs) and academia-industry cooperation are a must to mitigate risks in the development of AI systems. The accountability to mitigate these risks should not be with the developers only. The air commander must have the professional advice of the security professionals about the threats presented by adversarial AI.[41]Adversarial AI defence requires a 360 degree approach called the MLOps or AIOps, used interchangeably. This is the complete life-cycle, from design, development of algorithms and models, T&E, data collection, curation and validation to implementation under the assessed risk conditions.[42]

In an AI attack, the ability to generate enough diverse examples to thwart the stochastic process of the AI model is crucial. This is also the core to use synthetic and generative data to train by mimicking such an adversarial attack on own AI-based systems. Such a capability ensures ascertaining the robustness of own systems and creates a repository of examples to harden

40. Ibid.
41. Patrick Hinton, "Adversarial AI: Coming of Age or Overhyped?" Centre for Emerging Technology and Security, Alan Turing Institute, CETaS Expert Analysis (September 2023), https://cetas.turing.ac.uk/publications/adversarial-ai-coming-age-or-overhyped
42. Tom Olzak, "Adversarial AI: What It Is and How to Defend Against It?" Spiceworks AI. cybersecurity researcher, author and educator, June 28, 2022, https://www.spiceworks.com/tech/artificial-intelligence/articles/adversarial-ai-attack-tools-techniques/

own AI models.[43] The linkages of cyber security processes and responses with adverserial AI actions need to be examined with greater inquisitiveness by all government and private establishments in their AI deployment. The cyber security policies of the Computer Emergency Response Team (CERT) need to consider an attack on AI-assisted platforms as a national security intrusion and formulate proactive alarm systems for blocking the adversarial AI attack. Counter-AI tools have been proposed and developed that can assess the AI system's security before launching it on any platform. Tools can analyse the aspects of adversarial T&E, and the combined AI system red teaming process in various combat scenarios.[44] These tools need to be developed in-house, within India, with the collaborative knowledge resident with the academia, industry and air force. System architecture design, detailed system specifications, counter-AI tool user scenarios, user interfaces, etc. need to be formulated for AI-enabled air platforms. Such an exercise will also augment adversarial AI awareness among the air commanders.

Satellite imagery is gaining prominence with many governments as well as private players entering the race to provide low orbit satellite imagery in various Electro-Magnetic (EM) spectra. The cost of such information in real-time has been continuously coming down with the proliferation of many players across the globe. India has recently allowed the privatisation of satellite technology and is now open to manipulation by adversaries. These space platforms operate in a very hostile environment and their degradation against adversarial attacks increases the location uncertainty and limits their field of view, which have to be countered.[45] The effects of adversarial AI on missions using the satellite platforms' overhead imagery could be devastating for national security. AI models can analyse the imagery to

43. Dr Darminder Ghataoura, "Adversarial AI Fooling the Algorithm in the Age of Autonomy", Fujitsu Group White Paper Adversarial AI. https://www.fujitsu.com/uk/imagesgig5/7729-001-Adversarial-Whitepaper-v1.0.pdf
44. Nathan Byington, Carter Davis, Matthew Meehan, Caroline Vincent, David Woodward, and Nathaniel Bastian, "Counter-AI Tool System Design for AI System Adversarial Testing and Evaluation", Proceedings of the Annual General Donald R. Keith Memorial Conference, West Point, New York, USA,April 28, 2022, https://www.ieworldconference.org/content/WP2022/Papers/14-GDRKMCC-22.pdf
45. n. 19.

However, there is a need to encompass defence AI capabilities in the National AI Strategy, which presently includes only the commercial and private segments.

detect, identify, track and inform, and so can an adversary with the commercial satellite data sources. Adversarial AI can develop patterns of movement that could render traditional mathematical tracking approaches or maintaining track continuity across sensor gaps difficult or ambiguous even if a multi-modal spectrum is used by the satellite for imagery. Also, of concern are the several of points of attack along the information path, from the sensing of a physical phenomenon to the display of targets of interest, to mission personnel.[46] The generative AI and sensor-fused tracking solutions in such attacks are, therefore, crucial. Thus, on the satellite air platforms, it is critical to have AI standards, associated system engineering practices, expansive testing and evaluation of AI tools, techniques and procedures to understand adversarial AI risks, and alleviate consequent system threats to the mission success.[47]

CONCLUSION

AI is the technology which has proliferated in the civil domain and thereafter has been adopted by militaries across the globe. A similar trend is visible in India as well. The NITI (National Institution for Transforming India) Aayog released the national strategy on AI in 2018. In 2019, the Ministry of Defence (MoD) set up the Defence AI Council (DAIC) to provide strategic leadership for AI adoption in defence, and a Defence AI Project Agency (DAIPA) for enabling AI-based processes in the defence sector. However, there is a need to encompass defence AI capabilities in the National AI Strategy, which presently includes only the commercial and private segments. In the IAF, the AI Centre of Excellence has been formed at the Air Force Station, New Delhi, under the aegis of UDAAN (Unit for Digitisation, Automation, Artificial Intelligence and Application Networking) and the Indian Navy

46. Ibid.
47. Ibid.

made the INS *Valsura*, the centre of excellence in the field of AI and related technologies. With AI being employed in indigenous air warfare and other military platforms, the civil-military and PPPs have to be aggressively nudged by the government.[48] The edifice of AI incorporation in India has been laid and it must be ensured that it is safe, protected and abreast with ever changing technological AI innovations. Adversarial AI attacks have to be foiled to maintain trust in our systems, and, similarly, offensive AI capabilities must be developed to exploit the adversary's AI weaknesses. The collaborative approach of the military, domain experts, private sector, national laboratories and institutes cannot be over-emphasised for harnessing the strength of AI. These resources need to formed into action teams, thereafter nurtured, trained and directed to achieve the adversarial AI capabilities collaboratively and seamlessly in the civil and military domains. Seminars, brainstorming and workshops must be regularly conducted involving the defence industry, user services, software developers and academia involved in AI research.

The collaborative approach of the military, domain experts, private sector, national laboratories and institutes cannot be over-emphasised for harnessing the strength of AI.

"Autonomous weapons are already a clear and present danger, and will become more intelligent, nimble, lethal, and accessible at an unprecedented speed."[49] The very intelligence of the AI system is based on its training but the decisions are unpredictable though very confident when operating on data outside of training distribution. The air commander would like the system decision to be demonstrative of low confidence when operating in

48. Sanur Sharma, "Artificial Intelligence in Warfare", The BACKGROUND NOTE ON AI, (LARRDIS NO.MP-IDSA/15/2022 dated July 2022).Research and Information Division, Parliament Library and Reference, Research, Documentation and Information Service (LARRDIS), Lok SabhaSecretariat, New Delhi, https://parliamentlibraryindia.nic.in/lcwing/Artificial_Intelligence_in_Welfare.pdf
49. Kai-Fu Lee, *AI 2041: Ten Visions for Our Future* (Crown Publishing Group, 03/05/2024, ISBN-13: 9780593238318).

an environment not seen before, and "fail gracefully".[50] As of now, a really formidable defence algorithm that can defy a wide variety of adversarial attacks or even the other way around, has not been designed, which puts the issue of trust in the AI systems on a very weak footing,[51] making it a challenge for the air commanders' concerns about susceptibility to adversarial attacks, particularly in autonomous air platforms.

50. Anant Jain, "Breaking Neural Networks with Adversarial Attacks: Are the Machine Learning Models we Use Intrinsically Flawed?" Data Science, published in *Towards Data Science*, February 9, 2019, https://towardsdatascience.com/breaking-neural-networks-with-adversarial-attacks-f4290a9a45aa
51. Ibid.