Issue Brief

November 2026
No: 566

The Invisible Breach: Defending C2 Systems Against Adversarial AI and Cognitive Poisoning

Brig GS Klair

The Invisible Breach: Defending C2 Systems Against Adversarial AI and Cognitive Poisoning

Abstract

As the tri-services embed artificial intelligence into command and control, intelligence fusion, and logistics, a new vulnerability has opened alongside operational gains: the deliberate corruption of the AI systems themselves rather than the networks that carry them. These brief calls this 'cognitive poisoning'—the manipulation of training data, live inputs, or system outputs so that an AI system quietly steers a commander towards taking an incorrect decision while continuing to appear as though it is functioning normally. Drawing on India's Evaluating Trustworthy AI framework, the Draft Defence Acquisition Procedure 2026, and comparative practices from the United States and China, the brief argues that India has taken useful first steps but has not yet built the institutional muscle — a Red Team, a command-authority doctrine, and procurement standards written specifically for poisoning resistance– needed to close this gap before an adversary tests it in a live crisis.

Keywords: Cognitive Poisoning, Adversarial AI, C2, Data Poisoning, Prompt Injection, Defence Acquisition Procedure 2026 (DAP 2026), Cognitive Warfare, Cyber Security

Introduction

The Indian Armed Forces are rapidly integrating artificial intelligence into command systems, intelligence analysis, logistics management, and tactical operations. Notably, the growing reliance on AI carries a threat that is seldom discussed in strategic circles: cognitive poisoning or the deliberate manipulation of AI systems through adversarial methods.

At its core, cognitive poisoning is corruption aimed not at a network but at the judgement. A conventional cyber-attack targets servers, firewalls, and data storage. Cognitive poisoning goes after the analysis a commander actually reads before taking a call, which means an adversary who succeeds at it, can easily shape Indian military decisions for weeks or months before anyone notices anything amiss.

There are three broad methods that an adversary can pursue. The first is corrupting the training data itself, so the system absorbs false patterns from the outset. Another is injecting

malicious instructions during live operations, which pushes a functioning system towards generating false recommendations. A third, subtler route is manipulating how the system presents its output by suppressing critical warnings, or presenting biased conclusions as neutral analysis, without touching the underlying data at all.

Recent conflicts show this is not theoretical. In Ukraine's war with Russia, information operations combined traditional disinformation with AI-generated content to create misperceptions in military command, most visibly in March 2022, when hackers broadcasted a fabricated video of President Volodymyr Zelenskyy appearing to call on Ukrainian forces to surrender. The clip was identified and removed within hours but is still widely cited as an early marker of how quickly synthetic media can be inserted into a live conflict (de Abreu, C.M. 2022 and Allyn, B. 2022). As adversaries develop AI capabilities of their own, they will employ cognitive poisoning as part of their warfare strategy. China's own doctrinal writing on "intelligentized warfare" already treats the cognitive domain as a battlespace in its own right. Chinese military theorists have argued that dominating an adversary's perception and decision-making can deliver strategic results without direct engagement (Harper, J. 2022)

India's defence forces may not be adequately prepared to combat this threat. Pertinently, to partially counter this threat the Evaluating Trustworthy Artificial Intelligence (ETAI) framework was released jointly by the Chief of Defence Staff (CDS) and DRDO in October 2024, which organises trustworthy AI around five principles viz. reliability and robustness, safety and security, transparency, fairness, and privacy. At the launch, the CDS stressed that military AI systems must be resilient to attacks from adversaries (MoD, 2024). Procurement standards do not require vendors to demonstrate resistance to poisoning attacks. Furthermore, the doctrine offers no clear line on when a commander should verify an AI recommendation and when they can simply act on it. There is no standing Red Team tasked with testing deployed systems for this type of vulnerability. The training curriculum does not adequately include instructions on recognising AI-generated misinformation or maintaining command authority when AI systems may be compromised.

This brief recommends four immediate actions. **First**, establish an AI Red Team within the Headquarters Integrated Defence Staff with the authority to conduct regular adversarial testing of all deployed military AI systems. **Second**, develop military doctrine specifying command authority over AI in crisis operations, including clear procedures for when commanders must verify AI recommendations and when to trust them. **Third**, use the ongoing

revision of the Defence Acquisition Procedure, the Draft DAP 2026, released for stakeholder consultation in February 2026, as the vehicle to write poisoning-resistance and adversarial-robustness testing into acquisition standards for military AI systems, rather than treating this as a separate exercise to be taken up later (MoD, 2026). **Fourth**, include cognitive-resilience training into key courses of instruction for mid-level officers and above in order to enable commanders to learn to spot AI-generated misinformation and hold their own judgement when a system is under attack. None of this requires waiting for a crisis to unfold—the threat is moving quickly, but so far, it remains one India can still get ahead of.

Understanding Cognitive Poisoning: The Contested Domains of Modern Warfare

Military planners have traditionally organised their thinking around five domains of warfare: land, sea, air, space, and cyber. Sixth, the cognitive domain has already made its way onto that list. The cognitive domain is simply where commanders think and decide; it is the space of perception, information, and judgement that sits behind every order given. Winning it has nothing to do with better decisions in the abstract— it means making better decisions than the other side. Losing it is quieter and more dangerous because a commander who has lost the cognitive fight often does not know it until well after the fact.

Militaries have long used psychological operations, deception, and propaganda to influence the cognitive domain. What is new is the speed and scale at which artificial intelligence can now conduct cognitive operations. An AI system can process thousands of intelligence reports, identify patterns that humans would miss, and generate recommendations for commanders in seconds. This speed is a significant advantage. It is also an enormous vulnerability.

The Rise of AI in Military Command Systems

India's tri-services are deploying AI across all military domains. The Indian Army is experimenting with AI-assisted command systems and autonomous systems for reconnaissance. This is no longer purely experimental: the army's own AI roadmap now targets full integration of AI, machine learning and big-data analytics into decision-making, operational planning and command and control by 2027, with 2026 designated the “year of networking and data centricity”. The army itself has stated that elements of this capability, including smart war rooms and real-time battlefield surveillance, were already visible during

Operation Sindoor in May 2025 (Mohan, V. 2026). The Indian Air Force is integrating machine learning into air defence systems to process radar data and make targeting recommendations. The Indian Navy uses AI to optimise logistics networks and predict maintenance needs on ships. These systems represent genuine operational advantages. A system that can process intelligence from hundreds of sources faster and more reliably than a human analyst is of great military significance. However, vulnerability sits underneath the speed. What happens once an adversary finds a way into these systems, rather than around them? A recommendation that a commander trusts without a second thought could, in that scenario, be the product of data that an adversary planted months earlier.

What Cognitive Poisoning Actually Means

Cognitive poisoning is the deliberate corruption of an AI system in order to mess with the decisions it supports. This is different from conventional hacking or data theft. A hacker steals information or knocks a system offline, and the damage is visible almost immediately. Poisoning does neither—he system keeps running, nothing looks broken, and the only casualty is the quality of the information a commander is working from. It generally takes one of the three undermentioned forms.

Data Poisoning: This refers to feeding an AI system with false or misleading information during training. Train an imagery-analysis system on doctored satellite photographs and it will misread real ones later. Train a forecasting tool on falsified conflict data and its predictions will simply be wrong, confidently and consistently, not because anything is broken, but because it learned from corrupted information to begin with.

Prompt Injection: Here the attack happens live rather than at the training stage. An adversary slips false instructions into the data stream an operational system is reading in real time, and the system responds to what it has been told rather than to what is actually true. A manipulated sensor feed reporting a threat that does not exist will produce recommendations built entirely on that false premise.

Output Poisoning: The underlying data and processing may be entirely sound, but the presentation layer is not. A critical warning gets buried under low-priority noise, a biased conclusion is presented as a neutral analysis, or the options a commander is shown happen to be the ones that suit the adversary.

These three mechanisms are not unique to the military domain, nor are they hypothetical or unexamined. The US National Institute of Standards and Technology's Adversarial Machine Learning taxonomy, the closest to an internationally recognised reference document on this subject, formally catalogues poisoning, evasion, and prompt-injection attacks against both predictive and generative AI systems. MITRE's Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS) framework catalogues real-world adversarial AI incidents, as MITRE's Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK) framework catalogues conventional cyber intrusions (National Institute of Standards and Technology, 2025; MITRE, n.d.). What is specific to the military application is not the technique but the consequence: a poisoned recommendation from a shopping algorithm produces a bad product suggestion, whereas a poisoned recommendation about troop movement can result in catastrophe.

What makes cognitive poisoning different from propaganda or disinformation is that it targets decision support systems themselves, not just human perception. A commander trained to spot propaganda will usually spot it. However, few are trained to ask whether the AI system briefing them has itself been compromised.

Why This Matters for India

Consider a realistic scenario: a border crisis unfolding at multiple locations simultaneously. Indian military AI systems have been quietly poisoned in advance through corrupted intelligence feeds, not pervasive enough to trip an alarm, but consistently enough that every recommendation the system generates leans, in small ways, towards whatever serves the adversary.

Working against time, commanders lean on exactly the systems that have been compromised, and the strategic decisions that follow are shaped by recommendations that were never trustworthy to begin with. By the time anyone traces the problem back to poisoned data, the adversary already holds the initiative, and the crisis has moved in whatever direction the adversary intended.

None of these events need to occur. However, avoiding it depends on work done well before a crisis, not during one.

Speed is what makes this perilous, rather than merely theoretical. An air-defence commander deciding in seconds has no time to check a recommendation by hand. A logistics officer relying on optimisation software could build supply vulnerabilities into a plan without ever realising that the underlying data was compromised. A planner modelling an entire campaign based on AI output could be working from predictions that were poisoned from the start.

Precedent from Recent Conflicts

Ukraine offers the clearest real-world precedent so far. Russian information operations blended old-fashioned disinformation with AI-generated deepfakes and synthetic videos, and the clearest early instance came in March 2022, when hackers broke into a Ukrainian broadcaster's website and news ticker to run a fabricated video of President Zelenskyy calling on his own forces to surrender. The video was debunked and removed within hours, helped by the fact that Ukraine's government had pre-warned the public that such a fake video might appear and by Zelenskyy's own rapid rebuttal on Telegram. Digital forensics experts at the time nonetheless described it as only "the tip of the iceberg" for what AI-enabled information warfare would look like as the technology matured (Allyn, B. 2022). While the Ukrainian military leadership maintained its overall operational effectiveness, the conflict clearly demonstrated how AI-enabled information operations could create confusion in command chains and slow decision-making.

The sophistication of these operations will only increase. Any adversary that develops comparable AI capabilities against India would have every reason to fold cognitive poisoning into a wider information warfare campaign and not treat it as a standalone tool.

Vulnerabilities in Current Indian Military AI Systems: Speed versus Security

There is real pressure on the Indian defence forces to field AI systems quickly to keep pace with what potential adversaries are building. That pressure has a cost— a system rushed into service is often not rigorously tested against adversarial manipulation. Military acquisition processes typically include cybersecurity testing to ensure that systems have firewalls, encryption, and access controls. But cybersecurity and cognitive security are different things. A system can be cybersecure while still being cognitively vulnerable. It can have perfect encryption and still be poisoned.

Current military procurement standards do not systematically require the testing of AI systems against poisoning attacks. Vendors are not required to demonstrate that their systems can detect and resist corrupted data or malicious inputs. Security testing focuses on protecting the system from external attack, not on protecting the quality of the system's outputs. This is beginning to change on paper. The Ministry of Defence released the draft Defence Acquisition Procedure 2026 for stakeholder comments on 10 February 2026 and, unlike its predecessor, it carves out dedicated acquisition protocols for AI, cyber, and electronic-warfare systems. It introduces Technology Readiness Level (TRL) categorisation for emerging technologies requiring, for instance, TRL 5–9 certification for indigenously designed systems entering procurement. Furthermore, it shifts the underlying philosophy of acquisition from "Made in India" towards "Owned by India", retaining source code, critical design data, and upgrade authority rather than relying solely on the transfer of technology from foreign vendors (MoD, 2026).

Data Integrity and Training Quality

All AI systems depend on the quality of their training data. Training a system on correct data yields accurate results, whereas training it on corrupted data yields compromised outputs. Military AI systems typically draw on a wide mix of inputs, such as historical records, live intelligence databases, satellite imagery, signals intelligence, and occasionally open-source material, and each of those feeds is a separate door an adversary could try and exploit.

There is, at present, no standardised procedure across military institutions for:

- Verifying training data integrity before it is used to build a system.
- Detecting when training data has been corrupted or poisoned.
- Keeping a full provenance trail so that an error can be traced back to its source.
- Quarantining compromised data, sanitising and then retraining the system from a clean baseline.
- Securing the storage of sensitive training datasets against tampering.

In the absence of these safeguards, an adversary with any access to the data pipeline can introduce corruption subtle enough to go unnoticed for a long time. It might not surface at all

until a system starts producing a string of incorrect recommendations in the middle of a crisis, which is the worst possible moment to discover it.

The Supply Chain Issue

India's defence technology supply chain includes several participants. A military AI system may use code libraries, pre-trained models, and data-processing tools from different vendors, both local and international. Each component could be a potential failure point that users might not detect.

A foreign vendor might intentionally create a vulnerability to gather intelligence. However, a more common situation is that a vendor's supply chain is unknowingly compromised, allowing the issue to go unnoticed. Either way, a compromised third-party library can leave a backdoor or a poisoning capability sitting inside a system nobody thought to check that closely.

Routine maintenance is its own risk window. A contractor with access during an update has an opportunity to introduce a malicious change that looks, on the surface, like ordinary servicing. Current personnel-security screening is not obviously equipped to catch a sophisticated insider who understands AI systems well enough to hide what they are doing.

Existing vendor security standards ask for certifications but stop short of requiring proof of poisoning resistance or real transparency into the supply chain. Draft DAP 2026 aids with its push towards Indian ownership of source code and design data, thereby narrowing the supply chain simply by reducing how many components need to be trusted blindly. Ownership is not the same thing as verification, though. Nothing in the draft, as currently understood, requires anyone to confirm that the code or the data used to train whatever AI sits inside it is actually clean (MoD, 2026).

The Command Authority Challenge

Perhaps the most critical vulnerability is doctrinal. The current Indian Military Doctrine does not clearly specify when commanders should trust AI recommendations and when they should verify or override them. In crisis situations, where decisions must be made rapidly, commanders need clear guidance. Trust every recommendation without question, and you are exposed to poisoning. Verify everything by hand and you lose the speed AI was intended to provide in the first place.

Doctrine needs to draw that line, spelling out which categories of recommendation (strategic movement, targeting, logistics optimisation, and so on) demand verification before anyone acts on them, and which can reasonably be trusted with a lighter touch. It must explain the procedures to follow during the command of forces when AI systems are suspected to be compromised. It should guide commanders to recognise when AI output contradicts human expert judgement and how to respond.

Currently, no such doctrine exists in Indian military institutions. The ETAI framework's emphasis on resilience to adversarial attacks is a useful statement of principle. However, it is not an operational doctrine, as it does not inform an air-defence commander, in a specific instance, whether a targeting recommendation should be trusted or verified (MoD, 2024).

Risks in Intelligence Sharing

India shares intelligence with partner nations and receives it from them, and a good deal of that intelligence eventually feeds the Indian military AI systems. This aspect highlights the fact that coalition operations bring their own separate set of vulnerabilities into the picture.

A poisoned intelligence product shared across an allied network does not stay contained to one military's systems; it can corrupt several at once. The reverse is equally true: an Indian AI system compromised at the source could end up degrading decision-making in a partner country that never had a reason to distrust what it was receiving.

None of the protocols currently governing allied intelligence sharing were written with this risk in mind, and there is no agreed process for flagging suspected poisoning or alerting a partner quickly once it is found.

The Training and Awareness Gap

At present, military personnel, encompassing officers of every rank, lacks structured training to identify misinformation or synthetic content produced by AI. There is no clear and proper instructions for officers on critically evaluating AI system outputs or questioning the potential contamination of AI-generated recommendations.

With the growing dependence on AI in the military, the training gap is becoming increasingly critical. It is not surprising that a commander lacking education in identifying AI-generated disinformation is susceptible to it. A logistics officer with no grounding in supply-

chain poisoning is unlikely to notice a subtle anomaly buried in an AI-generated recommendation.

The Strategic Stakes for India

A Competitive Advantage That Could Evaporate

Much of the Indian military strategy assumes that AI will deliver a real operational edge over potential adversaries. This assumption only holds if the underlying systems are protected against poisoning. Left unprotected, the same advantage can just as easily flip against India.

An adversary that can poison Indian military AI systems while protecting its own would gain a decisive edge. This advantage would not come from superior technology or an edge in combat ratios. This would result from degraded decision-making on the Indian side and better decision-making on the adversary's side.

In multi-domain operations involving the Army, Navy, and Air Force, a poisoned AI system in one service could degrade coordination across all services. A corrupted feed reaching the Air Force could translate into poor air support for troops on the ground. A poisoned Army logistics system could quietly create supply gaps the Navy would be depending on without knowing it.

This threat extends beyond the battlefield. An opponent aware of the fact that Indian AI systems are susceptible to tampering might be motivated to exploit this vulnerability, not necessarily for military gain but to undermine political trust in military leadership. If AI-driven military decisions repeatedly lead to failures, they could result in political demands to restrict military actions or alter strategic plans.

The Cognitive Warfare Doctrine of Adversaries

Potential adversaries, most notably China, have developed explicit doctrines treating the cognitive domain as a warfare domain. The People's Liberation Army's writing on “intelligentized warfare” formally places the cognitive domain alongside land, sea, air, space, and cyber. Senior PLA theorists have argued that achieving what they call “mind dominance” over an adversary's decision-makers can deliver strategic results without direct engagement—a contemporary reading of the old idea of subduing an enemy without fighting. The Pentagon's yearly evaluation of China's military strength has independently highlighted Beijing's intention to employ AI to conduct large-scale influence campaigns, such as deploying automated content strategically to impact a specific audience effectively (“The Future of China's Cognitive Warfare,” 2022; DefenseScoop, 2022). Both strands of doctrine aim to influence the

information a decision-maker receives and the perceptions formed by the entire civilian population.

As these abilities develop, incorporating cognitive poisoning into strategic plans is a logical move for any opponent developing this type of AI capability, along with the necessary operational plans to implement it. Indian military leaders should assume that adversaries are already considering ways to poison the AI system used by the Indian defence forces.. The question is whether India is well poised to defend against these attacks.

Crisis Escalation Risk

The mechanics of this risk were set out earlier in this brief: an adversary who poisons Indian military AI systems ahead of a crisis does not need to trigger an obvious failure, only a systematic bias in the recommendations that commanders receive. What is worth adding here is where, in India's strategic environment, such a crisis is most likely to unfold. India faces the possibility of simultaneous friction on multiple fronts. Notably, a two-front scenario is precisely the setting in which commanders are under the greatest time pressure and are therefore most likely to lean on AI-generated recommendations without full verification. An adversary capable of poisoning Indian military AI would have every incentive to time that poisoning to coincide with exactly this kind of multi-front pressure, when the cost of pausing to verify a recommendation is highest and the temptation to trust it outright is greatest. This is not a reason to slow AI adoption; it is a reason to make sure verification doctrine and red-teaming are in place before, not after, such a moment arrives.

The Multi-Service Integration Challenge

The Indian defence forces are increasingly integrating operations across the Army, Navy, and Air Force. Integrating these systems is both essential and strategically wise; however, it introduces additional points of vulnerability. Each service crafts AI systems tailored to its specific functions. These separate systems need to interoperate during multi-domain operations.

If a poisoning attack damages an Army AI system that works with Navy AI systems, the damage could spread. A damaged Army logistics AI might provide suggestions that Navy supply chain AI systems use. A poisoned Air Force targeting AI could impact Army air defence

coordination. There is currently no standard protocol for handling cognitive security across systems involving more than one service.

Allied Partnerships and Shared Risk

India's expanding strategic partnerships bring real intelligence and operational benefits, thus giving the country access to allied intelligence and a foothold in coalition operations. However, these connections also pose a cognitive security risk: tainted allied intelligence could enter Indian systems without the ally realising it, and a compromised Indian analysis shared with allies could harm the relationship when it is most crucial.

An adversary who knows these connections might attack Indian AI systems because the harm could spread beyond India's borders. Protocols for handling cognitive security in shared intelligence among allies should be created in advance, not after problems arise.

The Domestic Technology Development Opportunity

India's defence research and development ecosystem includes the Defence Research and Development Organisation (DRDO) and private defence technology firms. These institutions are developing military AI systems for deployment by the tri-services.

India is strategically focused on establishing domestic military AI capabilities that are inherently resistant to poisoning attacks. This focus extends beyond mere technical preferences. This is a strategic necessity. Military systems developed by Indian engineers and understood by Indian institutions can be designed with cognitive security as a core principle rather than an afterthought.

Foreign-sourced AI systems may have vulnerabilities that are not obvious. Domestic systems developed in India and controlled by Indian institutions can be designed with transparency and security from the inception. Draft DAP 2026 already gestures at this ambition, describing India’s intent to retain source code and upgrade authority rather than simply hosting foreign manufacturing on Indian soil (Ministry of Defence, 2026). Cognitive security is a natural extension of that same logic— a system whose code, training data, and update pipeline are all under Indian control is inherently easier to audit for poisoning than one that is assembled from opaque foreign components. This is an argument for writing cognitive-security requirements into the domestic development ecosystem now, while it is still being built, rather than retrofitting them later.

How Other Militaries are Responding

India is not alone in confronting this problem and does not need to start from first principles. The United States' National Institute of Standards and Technology has published a formal taxonomy of adversarial machine learning covering poisoning, evasion, and prompt-injection attacks against both predictive and generative AI systems, which is increasingly used as a common reference vocabulary by the US Department of Defense and allied militaries, alongside MITRE's ATLAS framework, which catalogues real-world adversarial AI incidents (National Institute of Standards and Technology, 2025; MITRE, n.d.). Comparable thinking is beginning to appear in the defence planning of other Indo-Pacific states responding to the same Chinese doctrinal shift discussed above, with cognitive-domain resilience increasingly named as a distinct line of effort rather than folded silently into general cyber policy (National Institute for Defense Studies, 2022). None of these frameworks are military secrets. They are public documents, and adapting their vocabulary and testing methodology for Indian conditions would be far faster than building an equivalent framework from first principles. The AI Red Team recommended below would have an immediate, low-cost starting point by simply adopting the NIST and MITRE taxonomies as a baseline test suite and extending them to Indian-specific operational scenarios.

Policy Recommendations

This brief recommends four immediate policy actions to protect Indian military command systems from cognitive poisoning.

Recommendation I: Establish an AI Red Team

An AI Red Team must be established within the Headquarters Integrated Defence Staff with an explicit mandate to conduct adversarial testing of all deployed military AI systems.

Purpose and Scope. The AI Red Team should conduct regular poisoning attack simulations against live military AI systems to test their resilience to:

- Corrupted training data.
- Malicious input during crisis operations.
- Supply chain compromise.
- Intelligence source manipulation.

- Adversarial manipulation of AI output presentation.

The red team should develop test scenarios specific to Indian operational contexts, including:

- Border crisis scenarios.
- Multi-domain operations involving all three services.
- Coalition operations with allied partners.
- Counter-insurgency operations.
- Naval operations in contested waters.

The red team should produce:

- Annual assessments of vulnerabilities in deployed military AI systems.
- Recommendations for the remediation of identified vulnerabilities.
- Classified reports on detected attack methods and adversary capabilities.
- Updated defensive measures and detection protocols.

Composition. The red team could include:

- Officers from each military service with expertise in operations and AI.
- DRDO personnel with expertise in AI security and testing.
- Personnel from defence research institutes with AI expertise.
- Cleared private-sector AI security experts.
- Intelligence community personnel with expertise in adversary capabilities.

Implementation Timeline. Establish Red Team by January 2027. Begin operations by July 2027.

Budget. Much of the red team's early workload of building a baseline test suite out of the NIST and MITRE frameworks discussed above fits within existing R&D budgets. Bringing in clear private-sector AI security experts and establishing dedicated infrastructure, including isolated test environments and simulation tooling, is a different matter and will likely require its own timeline. The Ministry of Defence could plan for that cost upfront.

Recommendation II: Develop AI Command Authority Doctrine

The Headquarters Integrated Defence Staff should draft a doctrine that clearly sets out who holds decision authority over AI-generated recommendations in operational settings and crisis situations.

Content. At a minimum, the doctrine should set out:

- When a commander is required to verify an AI recommendation before acting on it.
- When time pressure justifies acting on a recommendation without the verification step.
- Separate, stricter protocols for high-consequence categories, such as targeting decisions, strategic movement, and intelligence assessments.
- How to issue orders when an AI system is suspected but not yet confirmed to be compromised.
- How to proceed when an AI suggestion opposes the decision made by a human expert in the field.
- A straightforward and unmistakable procedure for a commander to countermand an AI suggestion when required.

This doctrine is intended to function as an operational manual for the existing ETAI framework. Although the ETAI specifies that military AI systems need to be resistant to adversarial attacks, this doctrine should specify the steps a commander should follow when the robustness of such systems is in doubt (Ministry of Defence, 2024).

To integrate this doctrine into military training and operations, it should be incorporated into:

- Tri-Services operational procedures (SOP documents).
- Military training academy curricula at all officer levels.
- Command courses at the war colleges.
- Crisis management simulation exercises.
- Pre-deployment training for commanders.

Implementation Timeline. Commission drafting by February 2027. Complete draft by April 2027. Begin implementation in training academies by July 2027.

Budget. No additional funding is necessary. It should be included in the current doctrine development and training processes.

Recommendation III: Revise Defence Procurement Standards for AI Systems

The Ministry of Defence should use the current consultation on Draft DAP 2026 to update the defence procurement specifications. This update requires all military AI system vendors to show cognitive security standards as a clear part of TRL certification before obtaining procurement approval. The Draft DAP 2026 has already introduced TRL-based categorisation for emerging technologies, including AI. The change recommended here is to make poisoning-resistance and adversarial-robustness testing a named component of that certification, rather than leaving it to be inferred from general security-certification language that was written with conventional cyber threats in mind (Ministry of Defence, 2026).

As part of required security demonstrations, all vendors should provide evidence of:

- Supply chain security certifications documenting the integrity of all code libraries, pre-trained models, and components.
- Poisoning attack resistance testing conducted by independent security auditors.
- Model auditability: the ability to trace AI recommendations to specific training data and decision-making logic.
- Data integrity documentation: proof that the training data used to build the system has been protected and verified.
- An incident response plan for detecting and acting on suspected poisoning.
- Personnel security standards covering every vendor employee with access to the system.
- Full transparency on where training data came from and how it was verified.

To integrate these standards into procurement, they must be added to:

- Guidelines for the Defence Acquisition Procedure, particularly the criteria for TRL certification outlined in the draft DAP 2026.
- Request for Proposal documents for all military AI procurements.

- Vendor security agreements.
- Post-delivery maintenance and support contracts.
- Annual recertification requirements for deployed systems.

Implementation Timeline. Requirements need to be added to the draft DAP 2026. Subsequently, there is a need to develop detailed standards by April 2027. Implement for all new AI procurements by July 2027. Apply to existing systems by January 2028.

Budget. Incorporating these criteria into the DAP 2026 TRL certification language incurs no additional cost. Ongoing, independent poisoning-resistance audits of vendor systems are a separate matter and will require a modest recurring budget addition, as this is specialist testing that current procurement staff are not equipped to do in-house.

Recommendation IV: Build AI Cognitive Resilience Training

There is an inescapable requirement at every military training academy to build AI cognitive resilience into its curriculum and not leave it as an optional add-on.

Training Objectives. By the end of this training, personnel should be able to:

- Spot AI-generated misinformation, deepfakes, and synthetic content.
- Describe the manner in which a poisoning attack can reduce system performance while not completely shutting it down.
- Maintain assurance in decision-making and assert control, even if there is a suspicion that a system is under attack.
- Cross-check critical recommendations using alternative analytical methods.
- Promptly report a suspected compromise through the correct channel.
- Comprehend the implications of the supply chain associated with a military AI system for its security.
- Notice unusual patterns in the AI output that might indicate poisoning.

Role-Specific Training. The content should differ by role as highlighted below:

- **Senior Officers.** The strategic implications of cognitive poisoning, command authority protocols, decision-making in AI-compromised crises, and the risk to interservice coordination.
- **Operational Commanders.** The operational implications, how to verify AI recommendations, planning for degraded operations, and assessing system reliability in the field.
- **Staff Officers.** A working technical understanding of poisoning attacks, the indicators to watch for, what counts as a reportable anomaly, and how to support a commander attempting to verify a recommendation.
- **Intelligence Officers.** Assessment of intelligence source integrity, detection of poisoned intelligence feeds, and protocols for identifying compromised data.

Implementation Timeline. Develop curriculum by March 2027. Implement at all training academies by July 2027. Incorporate into all officer training academies by December 2027.

Budget. Developing the curriculum itself fits within existing training budgets. Rolling role-specific training out across every officer training academy is the more expensive part. It will need additional instructor hours and dedicated simulation infrastructure are required for exercises that are realistic enough to be useful. These costs require deliberate planning.

Conclusion

Tri-service command systems face an actual and increasing danger from cognitive poisoning— a threat that will intensify as defence forces increasingly rely on AI for decision-making support. This concern is not merely theoretical or crafted for a policy document; it is grounded in adversaries' proven abilities in information warfare, combined with evolving AI capabilities. Militaries worldwide have begun to recognise this threat. Waiting for painful lessons in actual conflicts to develop defensive responses is not a strategy.

The policy recommendations in this brief are designed to be implemented immediately and within existing institutional structures, building on foundations that already exist, such as the ETAI framework's statement of principle and Draft DAP 2026's new acquisition architecture for emerging technology. They do not require new military branches or massive fresh funding, although this brief has been explicit about the specific points at which some

additional, modest funding will genuinely be needed. What they require are policy decisions and implementation discipline.

The window for establishing a defensive posture is closing. As AI systems become more deeply integrated into military decision-making and adversaries develop more sophisticated poisoning capabilities, the cost of remediation will increase. Action can prevent future crises.

It would be prudent to regard cognitive poisoning with the same seriousness as a cyber-attack, information warfare, or a conventional military threat, and to let that judgement show up in policy, doctrine, training, and procurement standards alike. The invisible breach is defensible. It simply has to be taken seriously before, not after, it is tested.

Works Cited

Allyn, B. (2022, March 16). Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn. *NPR*. https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia.

Harper, J. (2022, November 30). Pentagon report warns that China may wield AI to enhance 'cognitive domain operations'. *Defense Scoop*.https://defensescoop.com/2022/11/30/pentagon-report-warns-that-china-may-wield-ai-to-enhance-cognitive-domain-operations/.

de Abreu, C.M. (2022, March 18). Debunking a deepfake video of Zelensky telling Ukrainians to surrender. *France 24*. https://www.france24.com/en/tv-shows/truth-or-fake/20220317-deepfake-video-of-zelensky-telling-ukrainians-to-surrender-debunked.

Ministry of Defence, Government of India. (2024, October 17). Framework and guidelines to integrate Trustworthy AI into critical defence operations unveiled [Press release]. Press Information Bureau. https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2065847.

Ministry of Defence, Government of India. (2026, February 10). Draft Defence Acquisition Procedure 2026 [Draft policy document, open for stakeholder consultation]. https://www.mod.gov.in/.

MITRE. (n.d.). Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS). Retrieved August 2026, from https://atlas.mitre.org.

National Institute for Defense Studies. (2022). PLA's intelligentized warfare: The politics on China's military strategy. NIDS Security Studies. https://www.nids.mod.go.jp/english/ publication/security/pdf/2022/01/05.pdf.

National Institute of Standards and Technology. (2025). Adversarial machine learning: A taxonomy and terminology of attacks and mitigations (NIST AI 100-2e2025). U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2025.pdf.

Outlook Business. (2026, February 18). DAP 2026 explained: From "Made in India" to "Owned by India" in defence procurement. https://www.outlookbusiness.com/deeptech/ decoding-the-draft- defence-acquisition-policy-whats-hot-whats-not.

The future of China's cognitive warfare: Lessons from the war in Ukraine. (2022). War on the Rocks. https://warontherocks.com/the-future-of-chinas-cognitive-warfare-lessons-from-the-war-in-ukraine/.

Tribune News Service. (2026, February 17). India AI Summit gives a glimpse of artificial intelligence systems in use by the Army. The Tribune. https://www.tribuneindia.com/news/india/india-ai-summit-gives-a-glimpse-of-artificial-intelligence-systems-in-use-by-the-army/.

CENTRE FOR LAND WARFARE STUDIES
CLAWS
अन्वेषणं ज्ञानस्य मुख्यम्।

About the Author

Brigadier GS Klair is a fourth-generation officer commissioned into the Armoured Corps. He is an alumnus of the National Defence Academy and Defence Services Staff College and has attended the Higher Air Command Course at the College of Air Warfare. The officer has served in varied terrains, handled capital procurements at the Headquarters Integrated Defence Staff and has been posted on a diplomatic assignment as the Assistant Defence Adviser in Sri Lanka. Presently, the officer is commanding an Armoured Brigade on the Western Front.

All Rights Reserved 2026, Centre for Land Warfare Studies (CLAWS)

No part of this publication may be reproduced, copied, archived, retained or transmitted through print, speech or electronic media without prior written approval from CLAWS. The views expressed and suggestions made in the article are solely of the author in his personal capacity and do not have any official endorsement. Attributability of the contents lies purely with author.