DAIO Defense AI Observatory

# Trusting Allied Machines

## Building an Asia-Pacific Defense AI Assurance Regime on Minilateral High-Trust Pipelines

Sanghyun Ahn

DAIO Study 26|37

Ein Projekt im Rahmen von

dtec.bw
Zentrum für Digitalisierungs- und Technologieforschung der Bundeswehr

Finanziert von der Europäischen Union
NextGenerationEU

**About the Defense AI Observatory**

The Defense AI Observatory (DAIO) at the Helmut Schmidt University in Hamburg monitors and analyzes the use of artificial intelligence by armed forces. DAIO comprises three interrelated work streams:

- Culture, concept development, and organizational transformation in the context of military innovation
- Current and future conflict pictures, conflict dynamics, and operational experience, especially related to the use of emerging technologies
- Defense industrial dynamics with a particular focus on the impact of emerging technologies on the nature and character of techno-industrial ecosystems

DAIO is an integral element of GhostPlay, a capability and technology development project for concept-driven and AI-enhanced defense decision-making in support of fast-paced defense operations. GhostPlay is funded by the Center for Digital and Technology Research of the German Bundeswehr (dtec.bw). dtec.bw is funded by the European Union – NextGenerationEU.

Ein Projekt im Rahmen von

dtec.bw
Zentrum für Digitalisierungs- und Technologieforschung der Bundeswehr

Finanziert von der Europäischen Union
NextGenerationEU

# Trusting Allied Machines

## Building an Asia-Pacific Defense AI Assurance Regime on Minilateral High-Trust Pipelines

Sanghyun Ahn

DAIO Study 26|37

Ein Projekt im Rahmen von

dtec.bw
Zentrum für Digitalisierungs- und Technologieforschung der Bundeswehr

Finanziert von der Europäischen Union
NextGenerationEU

**About the Author**
Sanghyun Ahn is a Republic of Korea Army lieutenant colonel specializing in defense AI security and risk management frameworks. He led the introduction of the risk management framework (K-RMF) in the ROK military, authored the ROK Defense AI Security Guidelines (2024), and has held command and policy positions in ROK military cybersecurity policy and technology. He writes regularly in the Korean press on defense AI assurance and policy, and presented an integrated K-RMF/AI RMF framework at NetSec-KR 2026. He completed his graduate studies in computer science in the United States. Sanghyun Ahn writes in his personal capacity. The views expressed in this study are the author's own and do not represent the position of the Republic of Korea Ministry of National Defense.

**Acknowledgments**
The author thanks Heiko Borchert for his support and encouragement. All remaining errors are the author's own.

**Design**
Almasy Information Design Thinking

**Imprint**
Sanghyun Ahn, Trusting Allied Machines. Building an Asia-Pacific Defense AI Assurance Regime on Minilateral High-Trust Pipelines. DAIO Study 26/37 (Hamburg: Defense AI Observatory, 2026).

Defense AI Observatory | Chair of Political Theory | Helmut Schmidt University
Holstenhofweg 85 | 22043 Hamburg | T +49 40 6541 2776
www.defenseai.eu | contact@defenseai.eu | @Defense_AIO

ISSN (online): 2749–5337
ISSN (print): 2749–5345

# Content

1 Summary ..... 6

2 The Problem: Sovereign Machines, Allied Wars ..... 10

3 Assurance Cannot Be Imported: The Four-Layer Framework ..... 14

4 A Region Without a Regime ..... 19

5 Minilateral High-Trust Pipelines as the Regime's Foundation ..... 23

6 The ASEAN Analogy and the Middle-Power Wager ..... 28

7 A Working Baseline and the Road to Mutual Recognition ..... 31
7.1 Why This Baseline ..... 32
7.2 Roadmap ..... 33
7.3 What Seoul Should Do ..... 35

8 Conclusion: Verified, Not Just Fast ..... 36

Literature ..... 39

# 1 Summary

Asia-Pacific allies are fielding sovereign defense AI (models they build and control themselves) while planning to fight together. Japan is deliberately cultivating SoftBank and Sakura Internet as pillars of a sovereign AI infrastructure even as it draws selectively on American systems where needed; South Korea is pushing defense AI transformation as a national agenda, while Naver and other major domestic firms enter the military market with their own foundation models; and the Philippines, wary of dependence on foreign partners and technology, is working toward a defense-AI capability of its own, though from a smaller industrial base than its neighbors. Taiwan shows the same turn in a sharper form. It is developing indigenous defense AI through its Intelligent National Defense Program, drawing on a strong civilian technology base. But its strategic position leaves it more dependent on the United States than any other actor in the region. Its defense establishment still struggles to align that ambition with operational need. The point here is not Taiwan's place in any ring. It is that the sovereign AI turn runs wider across the region than the Korea-Japan pair alone suggests, and that dependence and sovereignty pull against each other everywhere, not only in Seoul and Tokyo.[1]

Public debate about this divergence has fixated on ownership: who builds the model, who holds the data, and under whose flag it runs. But this study argues that assurance, meaning whether one ally can verify what another ally's machine will do, is the harder and more relevant question. The question is not hypothetical. When CSIS and Scale AI tested eight leading foundation models against 400 expert-designed international-relations scenarios, the models' recommendations diverged sharply: some were significantly more escalatory than others, and all showed some degree of country-specific bias.[2] Machines trained in different institutional and linguistic environments differ not only in performance but in judgment. One clarification is due. The benchmark this study draws on tested large language models. Much of the discussion here focuses on LLM-type systems too. But defense AI is not one method. As Heiko Borchert puts it, it is "a bag full of methods," and different methods fail in different ways and must be measured differently. A vision system that identifies a target needs benchmarks such as accuracy and false-positive rates, while a system that generates tactics needs measures such as exploitability, and no single method's benchmark can validate another's.[3]

1 Liao, *Intelligent National Defense Amid Strategic Ambiguity? Defense AI in Taiwan*. On Japan's sovereign-AI buildout around SoftBank and Sakura Internet, see Microsoft, "Microsoft Deepens Its Commitment to Japan."
2 Jensen et al., "Critical Foreign Policy Decisions (CFPD)-Benchmark: Measuring Diplomatic Preferences in Large Language Models."
3 Borchert, "The Very Long Game of Defense AI Adoption: Introduction," pp. 6, 21.

The regime proposed here must allow for this plurality from the start, even where its examples come from language models. How far the language-model findings above bind in practice also depends on how widely armed forces come to field such systems in the first place. One task lies beyond this paper. Building assurance benchmarks for each family of AI methods, not only for language models, is a large research agenda in its own right. This study does not try to close it. It marks the boundary of the present argument, and the natural step beyond it. Connecting such systems into a combined kill web[4] without common verification standards moves unverified risk at machine speed under the name of interoperability.

Importing assurance frameworks from a (hegemonic) supplier poses fundamental challenges. Assurance needs to combine evaluation-methodology research, measurement, third-party certification, and governance in one framework. So far, no single country, the United States included, supplies all four elements in an alliance-ready, mutually recognizable form for its partners. Alliance-ready means evidence partners can read without translation, certification partners can verify independently, and recognition arrangements that let one ally's evidence count in another's deployment decision. The last of these does not yet exist in any defense AI assurance regime. Washington's own trajectory confirms the diagnosis. National Security Presidential Memorandum 11 of June 2026 made "assurance" one of the four named pillars of US national-security AI policy, alongside adoption, adaptation, and accountability, requiring that national-security AI be designed to be reliable, robust, steerable, and controllable, and directing standardized Test, Evaluation, Verification, and Validation (TEVV) methodologies for national security AI. The AI Forge program launched by the Defense Advanced Research Projects Agency (DARPA), National Science Foundation (NSF), and NIST's Center for AI Standards and Innovation (NIST CAISI) has begun building the underlying evaluation science.[5] But what counts as a lawful order, a proportionate response, or a protected object varies with jurisdiction and doctrine. A benchmark built for one institutional context mismeasures a system operating in another. Allies can borrow one another's evaluation tools, but each tool originates in a particular legal and doctrinal framework, and its applicability in a partner nation must be assessed rather than assumed. What counts as a lawful, proportionate, or correct decision is anchored in each nation's law, doctrine, and strategic culture. This does not close cooperation off, but it does shape it. Allies need not merge their ground truths,

4 Dominguez, "Head of U.S. Military in South Korea Calls for 'Kill Web' Linking Seoul, Tokyo and Manila."
5 The White House, National Security Presidential Memorandum/NSPM-11, "Artificial Intelligence in the National Security Enterprise." On the AI Forge program, see DARPA, "AI Forge."

as long as they make them explicit: each can declare the standard its evaluation assumes, and exchange results on that basis. Trust here rests on being transparent about which evaluation standard was used, rather than on sharing a single one.

The Asia-Pacific is peculiarly exposed to this problem. Europe debates defense AI assurance inside a robust institutional framework: NATO's AI strategy and follow-on evaluation work, the EU's AI Act (for all its national-security carve-outs), dense standards bodies. The Asia-Pacific has no equivalent, and DAIO's own studies have documented the gap from the inside. Australia has named "Assurance of Autonomy" a public-good research activity and built a Defense AI Research Network (DAIRNet) with evaluation and testing as an explicit mission, but these remain national programs.[6] South Korea's defense AI institutions are fragmented across the ministry, the services, and the procurement agency.[7] Japan pursues a careful two-track strategy without a regional verification anchor.[8] Reuben Steff's study concludes that even New Zealand, the smallest of the Five Eyes, must build a dedicated TEVV and assurance function and make national caveats technically executable inside allied networks.[9] Every ally is building its own national assurance stack, and none can see inside its neighbor's. On current trajectories, the worst possible outcome is a region of mutually opaque verification regimes.

This study argues that to avoid the worst possible outcome, Asia-Pacific countries should use insights from regional verification regimes built for other purposes. For example, the region already operates one in the cyber domain. Time-sensitive cyber threat intelligence moves daily through standing institution-to-institution pipelines linking the United States, South Korea, Australia, Japan, and the Five Eyes partners. Allied cyber exercises such as Cyber Flag, in which Korea has participated annually since 2022, already rehearse the very collaborative practices an assurance regime requires: disclosing vulnerabilities to one another, jointly attributing campaigns, drilling shared response procedures. And the critical-and-emerging-technology standards consultations, including the AUKUS Pillar II adjacency under discussion with Seoul since the 2024 consultations, already exchange sensitive technical standards among the same countries. These minilateral high-trust pipelines are the natural foundation for a defense AI assurance regime. The core countries stay the same, the design logic can be adapted, while the evaluation metric needs to be developed.

The ASEAN Cybersecurity Cooperation Strategy demonstrates that a voluntary, non-binding, capacity-building framework can substantially improve information sharing across the region through a standing CERT network, codified threat-infor-

6 Layton, *Evolution not Revolution: Australia's Defence AI Pathway,* pp. 17, 21.
7 Park, *Will the One Ring Hold? Defense AI in South Korea*.
8 Tsuchiya, *Overcoming the Long Shadow of the Past: Defense AI in Japan*.
9 Steff, *Small Force, Big Data: Defense AI in New Zealand*, pp. 49–50.

mation exchange, and confidence-building measures, while honestly acknowledging capability gaps among its members.[10] Pacific Forum's Cyber ASEAN initiative shows regional actors designing their own assessment frameworks on principles of local ownership and agency.[11] These precedents underwrite the wager at the heart of the argument: assurance regimes need not be hegemonic projects. Middle powers, which can neither self-certify at scale nor compel others to accept their certificates, need mutual recognition most, and for that reason are structurally well placed to build such a regime.

Korea can move first. Korea's Risk Management Framework, or K-RMF, a NIST-family risk management framework adapted to Korean defense conditions, has been adopted across the ROK military, with application being phased in, and the military issued internal Defense AI Security Guidelines in 2024. Korea's experience integrating the trustworthiness functions of the NIST AI RMF into K-RMF is proof of concept that a working baseline can be built and operated. The regional baseline itself should be drafted jointly, in a NIST grammar that partners can read without translation, first among them the United States, whose Department of Defense issued a guide tailoring its cybersecurity risk management framework to AI systems in July 2025. What does "a NIST grammar" mean here? It means a shared vocabulary for reporting assurance evidence, so that one partner can read another's results without translation. It does not mean one imposed evaluation method. This distinction matters, because different AI methods keep their own benchmarks, and each nation fills the shared fields with its own criteria. Borrowing a common vocabulary does not mean ceding one's own standards to it.[12]

The following chapters develop the study's argument for an Asia-Pacific AI assurance regime step by step. The next section sets the core challenge of verifying allied AI solutions in the geostrategic context. Section 3 decomposes assurance into a four-layer framework and shows why no ally can import the whole stack from a hegemonic supplier, including the deeper problem that evaluation's ground truth varies with institutions. Then the study surveys the Asia-Pacific's missing regime as a gap documented from within DAIO's own country studies. Section 5 presents the central argument that the region already operates high-trust minilateral pipelines in the cyber domain, and they are the natural basis of an assurance regime. Section 6 discusses how ASEAN cyber cooperation serves as a workable precedent that can help advance a regional defense AI assurance framework, and how this tests middle-power agency. Section 7 proposes a concrete working baseline, jointly drafted in the NIST grammar and informed by Korea's K-RMF integration experience, and a phased roadmap. Section 8 concludes with recommendations.

10 ASEAN, *ASEAN Cybersecurity Cooperation Strategy 2021–2025*.
11 Manantan, *Cyber ASEAN: Advancing Cyber Resiliency and Capacity in Southeast Asia*.
12 US Department of Defense, *Artificial Intelligence Cybersecurity Risk Management Tailoring Guide*. On K-RMF and the ROK Defense AI Security Guidelines, see section 7.

# 2 The Problem: Sovereign Machines, Allied Wars

Two developments are accelerating simultaneously in the Asia-Pacific, and they are pulling in opposite directions.[13] The first is the sovereign turn in defense AI. The other is the deepening integration of allied operations. Each is individually rational, but together they create a trust problem no ally has yet built the mechanisms to solve.

The sovereign turn shows most clearly in procurement decisions. If every US ally simply adopted US digital technology (Palantir's Maven Smart System, for example), coalition integration would be an engineering problem. But reality is more complicated. Japan has chosen a calculated two-track strategy by using US systems where the alliance requires them, while deliberately building SoftBank and Sakura Internet into pillars of sovereign AI infrastructure. The Microsoft and OpenAI investment agreements show that Tokyo is not excluding the domestication of American capability, and Fujitsu's expanded Palantir cooperation, including a new Palantir AI Platform licensing agreement in August 2025, likewise targets the corporate sector.[14] That may now be reaching the command level. In June 2026 Tokyo confirmed it is weighing Maven as a leading candidate for the Self-Defense Forces' command and control (C2) system, while pursuing a domestic 'control tower' to arbitrate among AI platforms and limit dependence on any single foreign supplier.[15] Korea leans further toward independence. Amid the national push for defense AI transformation, Naver has launched a dedicated defense AI unit on top of its own foundation model, and other major technology and defense firms have joined the race.[16] The Philippines, wary of dependence on foreign partners and foreign technology, strives to build indigenous capabilities for defense AI.[17] Australia has designated AI, machine learning, and algorithms a sovereign industrial capability priority, imposing on itself national control of the technology and intellectual property of its defense AI.[18]

Growing sovereignty concerns were confirmed in June 2026. Under a US export-control directive (received on 12 June, three days after public release), Anthropic had to cut off all foreign access, allied nationals included, to its newest frontier models, Fable 5 and Mythos 5, within hours of the directive's arrival.[19]

---

13 For the global picture of this acceleration, and of its ambivalence, see Borchert, "Ambivalent Acceleration: War, Battlefield Deployment, and Money Push Defense AI But Fundamental Questions Remain Unresolved." Its findings – dominant emulation, AI shaped by legacy platforms and service verticals, and sharpening digital-sovereignty concerns – describe precisely the terrain in which this study's assurance argument operates. This paper updates and extends the findings of Borchert/Schütz/Verbovszky (eds.), *The Very Long Game*.

14 Microsoft, "Microsoft Deepens Its Commitment to Japan"; Fujitsu, "Fujitsu Signs New Licensing Agreement with Palantir"; "OpenAI and SoftBank Group Partner to Develop and Market Advanced Enterprise AI."

15 Jung, "Japan's Self-Defense Forces Adopt Palantir AI for Command"; Chiba/Nemoto, "Japan Weighs Use of Palantir, Anduril AI Systems for Defense Forces."

16 "AI Model Rivalry Expands into Defense Sector"; "South Korea's Naver, KAI to Develop Defense AI Platform"; "Naver Targets Defense AI Market."

17 Sygaco, *Reimagining Horizons: Defense AI in the Philippines* (forthcoming).

18 Layton, *Evolution not Revolution*, p. 18.

19 Anthropic, "Statement on the US Government Directive to Suspend Access to Fable 5 and Mythos 5"; "Anthropic Says Trump Admin Has Lifted Export Controls on Claude Fable 5 and Mythos 5"; Borchert, "Ambivalent Acceleration," p. 10.

Access was renegotiated and restored weeks later, but the precedent could not be unmade. Reuben Steff's New Zealand study treats the episode as a defining lesson in how conditional allied technology access can be.[20] For planners in Seoul, Tokyo, and Canberra the implication is identical: a model you do not control is capability that can be recalled, and sovereign AI is a rational hedge.[21]

But in parallel to growing sovereignty demands, military integration is advancing, too. At the September 2025 Freedom Edge trilateral exercise, the three countries practiced multi-domain operations across sea, air, and cyber on top of the now-permanent real-time missile-warning data feed. Time-critical data exchange between allied systems is already routine, and the US Pacific Multi-Domain Training and Experimentation Capability is bringing AI-enabled experimentation into the same exercise architecture.[22] The US Army's Project Convergence, the Navy's Project Overmatch, and the Five Eyes' Project Arcadia are building the shared technical infrastructure of AI-enabled coalition command and control (C2), and partners as small as New Zealand are already inside it.[23] Plan ANZAC, signed in March 2026, commits Australia and New Zealand to operating as an integrated force by 2035.[24] Combined operations on the Korean Peninsula have always presumed deep technical interoperability. Today an AI layer is being laid on top of it.

Sovereignty and integration demands come to a head in coalition doctrine, which assumes that once allied systems are connected, their outputs can be trusted, that is, a track fused by a partner's sensor model, or a target candidate surfaced by a partner's decision-support tool, is as reliable as one's own. Sovereign AI breaks that assumption, because independently developed models are not interchangeable instruments. The clearest evidence is the Critical Foreign Policy Decision (CFPD) benchmark developed by the Center for Strategic and International Studies (CSIS) and Scale AI. For the evaluation, 400 expert-designed scenarios spanning escalation, intervention, cooperation, and alliance dynamics, were run against eight major foundation models (Llama 3.1 at 8B and 70B, GPT-4o, Gemini 1.5 Pro-002, Mixtral 8x22B, Claude 3.5 Sonnet, DeepSeek V3, and Qwen2 72B).[25] Presented with the same crises, the models' recommendations diverged widely. Four of the eight were significantly more escalatory than the most restrained models, and all showed some degree of country-specific bias. These are general-purpose commercial models under laboratory conditions. Defense-specialized models trained on national doctrine and national data can be expected to diverge further.

20 Steff, *Small Force, Big Data*, pp. 7–8.
21 Chavez, Chilukuri, and Scanlon, "Sovereign AI Index."
22 "Trilateral Freedom Edge Exercise Wraps Off South Korea"; "US, South Korea and Japan Open Joint Air and Naval Exercise"; "PMTEC Pioneers AI-Enabled Warfare Capabilities."
23 Steff, *Small Force, Big Data*, pp. 21, 47.
24 Ibid., pp. 20–21.
25 Jensen et al., "CFPD Benchmark."

In addition, a March 2026 joint analysis by the UK National Cyber Security Centre and the AI Security Institute reported that a frontier model autonomously completed over half of a simulated 32-step enterprise network intrusion that would take a human expert roughly fourteen hours, at a computer cost of about £65 per attempt.[26] Battlefield reporting from Ukraine describes Russia-linked groups embedding language models in malware to generate system commands in real time.[27] Both examples illustrate that unverified systems also offer new attack vectors to adversaries. Furthermore, the controversy over AI-assisted targeting in Gaza warned every watching military: in any country that allied machines will connect, accountability falls on the commander who approved the machine's output. Thus, without precautionary measures in place, coalition partners will also share the reputational damage of one another's AI failures.[28]

Consequently, armed forces in the Asia-Pacific region (and elsewhere) need to ask themselves: Can Seoul establish, on evidence rather than assurances, how Tokyo's target-support model degrades in fog? Can Canberra know where Korea's sensor-fusion system resists deception and where it is vulnerable? Can a US commander connecting a partner's model into a combined kill web read that model's limits the way they read a coalition pilot's flight hours? Today the answer is no. The technical capacity for verification exists. What is missing is any regional structure through which allies produce, exchange, and mutually recognize verification evidence.

For Korea, these questions are ever more pertinent given recent legislative developments. In February 2026, lawmakers from both parties introduced the National Assembly's first bill on military AI, aimed squarely at the fact that Korea's landmark AI Framework Act, in force since January 2026, explicitly excludes the defense sector from its scope.[29] The bill sets basic policy directions for defense AI and requires human intervention where automated decisions could pose grave risks to life and bodily safety. Like most first-generation defense AI legislation, it is stronger on adoption than on assurance, a problem best corrected by integrating trustworthiness verification into the military's existing risk management framework rather than bolting it on afterward. But even the best-designed domestic statute will be limited as its reach is only national. Thus, the risk is that different allies will develop diverging, potentially incompatible, assurance regimes. And this will neither advance military integration nor sovereignty.

26 UK National Cyber Security Centre, "Why Cyber Defenders Need to Be Ready for Frontier AI."
27 On LLM-embedded malware in Ukraine, see: "Ukraine Says Russia Using AI Malware on Battlefield."
28 On AI-assisted target generation in Gaza, see: Abraham, "'Lavender': The AI Machine Directing Israel's Bombing Spree in Gaza."
29 "Bipartisan Military AI Bill Proposed at National Assembly"; Future of Privacy Forum, "South Korea's New AI Framework Act," 2025.

# 3 Assurance Cannot Be Imported: The Four-Layer Framework

Despite the need for assurance, the concept risks remaining a slogan everyone endorses, and no one specifies. Assurance is the production of transferable evidence about an AI system's behavior: where it can be trusted, where it fails, how it can be attacked, and under what conditions its outputs should be discounted. Producing such evidence requires four distinct layers of capability, each with its own institutions, techniques, and failure modes. That's why "buy American and the assurance comes with it" is not enough. Rather, Asia-Pacific countries need to purposefully build the four-layered framework:

- **Layer 1: Evaluation-methodology research**
  The science of evaluation and measurement provides the foundation to develop adequate adversarial robustness testing, interpretability, red-team methodology, benchmark design, and uncertainty quantification. The research base at this layer is thin everywhere and concentrated in a handful of institutions. The UK AI Security Institute, reorganized and renamed around security in February 2025, has become one of the leading evaluation bodies for frontier models, and the international network of AI safety institutes agreed at the May 2024 Seoul summit spans the UK, the United States, Japan, Singapore, Korea, and Australia, among others.[30] In the United States, AI Forge, the joint DARPA, NSF, and NIST CAISI program launched in June 2026, has committed national resources to interpretability, control, and adversarial robustness for national security AI and to credible, verifiable benchmarks.[31] For the time being, only a handful of countries build this foundational layer. Learning from these countries and bringing in one's own knowledge and requirements is the task Asia-Pacific countries should tackle jointly by cooperating on building a common foundational layer.

- **Layer 2: Measurement**
  The second layer applies the methodology to real systems to run evaluations and generate the trustworthiness data. This is where measurement creates evidence: A computer-vision model that achieves 99 percent target-recognition accuracy on a sunlit training range tells a commander almost nothing. What matters is how accuracy decays in fog and at night, how the model responds to camouflage and deception, how confidence fluctuates with sensor data quality. Measurement converts a system's strengths and limits into data a commander can read. A human who approves a machine's recommendation without knowing where the machine is reliable and where it is fragile cannot meaningfully be said to exercise control; the approval is closer to a formality.[32] Trustworthiness data is what makes human control real, letting people actually judge and intervene rather than formally approve, and it is what allies would exchange under the regime proposed here.

30 UK Department for Science, Innovation and Technology, "Tackling AI Security Risks to Unleash Growth and Deliver Plan for Change"; Seoul Declaration, AI Seoul Summit, 2024.
31 For the respective program materials, see: https://www.darpa.mil/research/programs/ai-forge (last accessed 14 August 2026).
32 Steff, Small Force, Big Data, p. 44.

- **Layer 3: Third-Party Certification**
  When the developer who built a system or the military that wants to field it measures it, an interest in favorable results is built in. Wishful thinking, however, undermines assurance. Independent evaluation to assess AI is available in the Asia-Pacific region, but it must be structured to meet defense requirements and cross-border recognition. For example, the AI Verify ecosystem developed under Singapore's Infocomm Media Development Authority pioneered voluntary AI testing against a published framework[33], and Resaro, a commercial assurance firm co-headquartered in Singapore and Munich, performs independent testing and evaluation of mission-critical and dual-use AI systems for public-sector clients.[34] As Peter Layton argued, Australia's Trusted Autonomous Systems center operates "assurance of autonomy" as one of two designated public-good activities, and DAIRNet exists explicitly to support the evaluation, testing, and integration of AI systems.[35] Early moves into military-grade evaluation, however, are so far mostly US-led. The region's own assets, described above, are still civilian and dual-use, or defense-oriented but national. The elements therefore exist, but they are not yet combined into a defense-grade system. Above all, mutual recognition is missing, that is, the agreement that a certificate issued in one capital counts as evidence in another. Building this recognition layer is a task the countries in the region should tackle jointly.

- **Layer 4: Governance**
  The final governance layer defines who sets baselines, who accredits certifiers, how deployment decisions incorporate assurance data, and who is accountable when a verified system fails anyway. This is where risk management frameworks operate (the US federal RMF family and its DoD variant, and Korea's K-RMF), and where the DoD's AI cybersecurity tailoring guide (July 2025) and NIST's draft Cyber AI Profile (NIST IR 8596) mark current, still-evolving American practice.[36] Governance is also the layer where legal and constitutional differences between allies collide hardest, which leads to the deeper reason assurance cannot be imported. That's why defining the governance layer will require dedicated efforts by the countries in the Asia-Pacific region to align expectations regarding the goals of the framework and the requirements that need to be met to comply with it.

33 On Singapore's defense-AI trajectory and its AI Verify ecosystem, see: Raska, "Reimagining Defense Innovation: Defense AI in Singapore," pp. 550-580.
34 Infocomm Media Development Authority (Singapore), AI Verify framework documentation; Resaro, "Resaro Levels Up: Establishing Co-Headquarters in Europe."
35 Layton, *Evolution not Revolution*, pp. 17, 21.
36 US Department of Defense, *AI Cybersecurity Risk Management Tailoring Guide*; NIST, *Cyber AI Profile*.

This four-layered framework rests on a fundamental question: How do you establish the ground truth for an AI solution to be assessed? The CFPD benchmark provides a partial answer in the form of identical scenarios, diverging recommendations and measurable country bias.[37] But the drivers that determine which scenarios are chosen for benchmarking are likely to vary from country to country, reflecting diverging strategic cultures and constitutional frameworks. What counts as a lawful order differs across jurisdictions. An instruction legally binding in one military can be one a partner officer is legally obliged to refuse. What counts as a proportionate response, a protected object, a permissible target of opportunity depends on national rules of engagement, treaty obligations, and constitutional structure. A decision-support model scored "safe" against one country's evaluation standard may be unsafe, or unlawfully conservative, against another's. Nor is ground truth static. National law, precedent, and doctrine evolve, sometimes abruptly, and when they do, the evaluation standard for decision-support AI changes with them. An evaluation regime must therefore be able to re-encode its ground truth based on user, not supplier requirements.

Language creates another challenge. The global evaluation apparatus (benchmarks, red-team corpora, safety test suites) is overwhelmingly English-based, while the defense AI fielded in the Asia-Pacific region operates in Korean, Japanese, and many other regional languages, on doctrine written in those languages, against adversary communications in yet others. Robustness measured in English is weak evidence of robustness in Korean. This matters as models' ability to refuse harmful requests is reported to weaken unevenly across languages.[38] One study found that merely translating jailbreak prompts into low-resource languages drove bypass rates from under 1 percent to 79 percent.[39] These are findings about general-purpose models rather than fielded military systems, but the message is hard to ignore. Simple translation of a benchmark does not carry the institutional and cultural context that gives scenarios their meaning.[40]

In the Asia-Pacific, a regional assurance regime that did nothing but build validated non-English defense evaluation corpora would already justify its existence. The implication is not that allies should go their separate ways, but that convergence must be built not only assumed. Each ally must maintain sovereign ground truth for the judgments its own law and doctrine define. What the alliance needs is a mechanism for exchanging evaluation results with their ground-truth premises declared, letting partners see not only what a system scored but whose evaluation standard it was scored against. That mechanism constitutes the core of trust-build-

37 Jensen et al., "CFPD Benchmark."
38 On multilingual safety evaluations reporting lower refusal rates for harmful requests in languages other than English, see, among others, Yong et al., "Low-Resource Languages Jailbreak GPT-4".
39 Yong/Menghini/Bach, "Low-Resource Languages Jailbreak GPT-4."
40 Ibid.; Wang, "All Languages Matter: On the Multilingual Safety of Large Language Models."

ing the proposed framework is meant to achieve. Making such premises explicit is itself a documented discipline. Value-based engineering standards such as ISO/IEC/IEEE 24748-7000 set out a process for eliciting stakeholder values and keeping them traceable through design, which shows that ground-truth premises can be declared and exchanged in a structured form rather than left implicit.[41]

If the four-layer analysis is right, its logic is not confined to defense. The same problem should appear wherever institutions handle consequential AI. Korea's Anti-Corruption and Civil Rights Commission, for example, has spent a decade transferring integrity-assessment systems to fourteen developing countries in partnership with UNDP. It used the partnership's tenth-anniversary forum in July 2026 to name AI and digital government the next frontier of integrity governance, with the warning that for decisions a system cannot explain, the humans and institutions that deployed it bear responsibility.[42] That is the measurement-and-governance logic that the proposed framework shall provide, but reached independently by anti-corruption governance. This example provides an early instance of the pattern advocated here, in which a middle power supplies a governance baseline that others localize. Assurance, in short, is the general form of the trust problem facing any institution that delegates judgment to machines; the defense version simply carries the highest cost of failure, which is why it should be the first to solve that problem institutionally.

41 ISO/IEC/IEEE 24748-7000:2022, Systems and Software Engineering — Life Cycle Management — Part 7000: Standard Model Process for Addressing Ethical Concerns during System Design (Geneva: ISO/IEC/IEEE, 2022).

42 "S. Korea, UNDP Look to Next Phase of Anti-Corruption Work as AI Risks Rise"; "Experts Warn AI Could Both Fight and Conceal Corruption"; UNDP Seoul Policy Centre, "'Korea-UNDP International Anti-Corruption Forum' Charts Future Global Anti-Corruption Strategies."

# 4 A Region Without a Regime

Europe argues about defense AI assurance inside a dense institutional network. NATO's AI strategy committed allies to responsible-use principles and to follow-on work on testing and evaluation, including through its Data and AI Review Board. The EU's AI Act creates a common legal framework for AI risk governance, although its application to defense and national-security activities is limited by significant carve-outs; and European standards bodies offer existing forums any assurance initiative can climb onto.[43] Whether Europe's institutions are producing assurance fast enough can be debated; that the institutions exist cannot.

In the Asia-Pacific they are absent. There is no regional NATO, no regional AI Act, and no standing body in which the region's US allies and partners could negotiate a common defense AI baseline even if every one of them wanted to start tomorrow. The region's top-level security arrangement is a set of hub-and-spoke bilateral alliances, supplemented by minilaterals of recent vintage. Its top-level multilateral bodies (ASEAN and its offshoots) are consensus organizations spanning the full spectrum of regional alignment, including states whose systems and suppliers are the very object of allied threat models. The institutional setup that makes a European-style assurance regime imaginable simply does not exist here.

This gap is even more relevant as all studies covering Asia-Pacific that the Defense AI Observatory has published since 2022 show that each state is building assurance dutifully at home, while state to state the systems remain mutually sealed.

Australia comes closest to a national assurance framework as envisaged in this study. Peter Layton discussed the Trusted Autonomous Systems cooperative research center pursuing "assurance of autonomy" as a designated public-good activity. DAIRNet supports the evaluation, testing, and integration of AI systems. The Royal Australian Navy's RAS-AI strategy names Sovereign Control, a control regime protecting national data, among the five fundamental effects it expects of autonomous systems, while its training doctrine aims to produce operators who trust but verify the technology.[44] Every element of the framework appears, but only as a national program. None of it is designed to make Australian assurance evidence readable in Seoul or Tokyo.

Youngwook Park's South Korea study presents the same problem in mirror image. Genuine AI development ambition flows through a fragmented institutional structure, with responsibility split across the ministry, the services, and the procurement agency, and a restrictive data regime narrowing what can be built and

43 NATO, "Summary of the NATO Artificial Intelligence Strategy," 22 October 2021, and subsequent Data and AI Review Board documentation; European Union, Regulation (EU) 2024/1689 (Artificial Intelligence Act), Official Journal of the European Union, 12 July 2024.
44 Layton, *Evolution not Revolution*, pp. 12, 17, 21, 39–40.

tested.[45] The defense AI bill now pending in the National Assembly would begin consolidating authority for adoption, but not yet its assurance counterpart. Korea's assurance assets, K-RMF and the 2024 Defense AI Security Guidelines, exist, as section 7 details, as national instruments without a regional interface.

Motohiro Tsuchiya portrays Japan as a state emerging from a long institutional shadow, building defense AI capability through a careful two-track arrangement with the United States and domestic industry.[46] That tension is no longer hypothetical. As section 2 notes, Tokyo is weighing the US-built Maven as a leading candidate for its command-and-control system while standing up a domestic 'control tower' to arbitrate among AI platforms and limit its dependence on any single foreign supplier. This deliberate mix of adopted and sovereign systems is exactly what allies would need to verify across the two-track seam, and no regional mechanism yet lets them.

According to Reuben Steff, New Zealand is told to build at the national level exactly what the region as a whole still lacks: a dedicated TEVV and assurance function able to explain not only what a system is meant to do but how it fails and how it can be attacked, and national caveats made technically executable inside coalition architectures through data tagging, auditable flows, and visible human approval points.[47] Steff's scenario analysis also makes it clear that a (small) state that cannot verify the AI systems it uses likely ends up with its decisions being shaped by others. This concern is also present in Indonesia, as Emanuel Ario Bimo and Bagus Jatmiko write in their study, which presents the country's approach as cautious and capacity-constrained, but fully aware of the risks of digital dependence.[48]

Against this background it might seem logical that ASEAN, the region's only standing multilateral body, should house the regime to come. As section 6 argues, ASEAN can play a helpful role, but for structural reasons it cannot provide the framework. Defense AI assurance evidence is alliance-sensitive material, and ASEAN's membership spans the full spectrum of regional alignment, including states whose systems and suppliers are themselves the object of allied threat models. As the next section will discuss, the regime must therefore start minilateral and remain compatible with ASEAN outreach, a two-speed design the region's security architecture already uses everywhere else.

Before proceeding, two alternative options should be addressed. The first is waiting for Washington to provide the respective framework. This creates a chal-

45 Park, *Will the One Ring Hold?*
46 Tsuchiya, *Overcoming the Long Shadow of the Past*.
47 Steff, *Small Force, Big Data*, pp. 44, 49–50.
48 Bimo/Jatmiko, *A Careful Walk on Thin Ice: Defense AI in Indonesia*.

lenge as NSPM-11 makes assurance a pillar only within the US national security apparatus. Nothing in the memorandum, and nothing constitutionally, supplies the mutual-recognition mechanisms that would make a Korean certificate count as evidence in Canberra, nor can it. A regional regime, by contrast, does not compete with American standards. It seems to be the only way through which allies can use them. Second, it could be argued that Asia-Pacific nations are not able to build such a regime, because they lack institutionalized cooperation mechanisms and regional membership is too heterogeneous. This argument is wrong. As the next chapters will show, the region already operates one high-trust technical exchange regime among exactly the right states, and it has already proven, in the ASEAN context, that soft regional frameworks can achieve institutional substance.

# 5 Minilateral High-Trust Pipelines as the Regime's Foundation

The region's US-aligned states already do, in cyberspace and daily, most of what a defense AI assurance regime would ask of them.

First, exchange sensitive technical evidence at speed. Cyber threat intelligence on indicators of compromise, malware signatures, attribution assessments of state-backed campaigns moves through standing institution-to-institution links that connect the United States, South Korea, Japan, Australia, and the wider Five Eyes community. Joint advisories co-signed by multiple national cyber agencies have become a routine instrument. An assurance regime for defense AI would perform the same tasks but focus on different subjects replacing indicators of compromise with evaluation results, red team findings instead of malware signatures, and model failure reports instead of incident reports. The trust threshold involved has already been crossed, since sharing what your cyber defenses missed is no less sensitive than sharing where your AI model is wrong.

Second, they train together under conditions that simulate real attack. Korea has participated annually since 2022 in Cyber Flag, the capstone allied cyber exercise, alongside Australia and Five Eyes partners. Such exercises routinize the collaborative practices an assurance regime requires: disclosing your vulnerabilities to partners, testing one another's defenses, and building shared insights from shared failures.[49] Exercises are also the single most natural venue for allied AI assurance to begin, for reasons the region has already demonstrated. Freedom Edge 2025 linked three militaries across sea, air, and cyber on top of a permanent real-time data feed, meaning allied digital systems are already being connected in exercise settings, with no assurance protocol evident in public reporting.[50] The precedent of exercises as technology-integration testbeds is equally established at the coalition's top end: Project Convergence and Project Overmatch are exposing partners to AI-enabled command-and-control experimentation at a scale no regional state could replicate alone.[51]

Finally, they already negotiate sensitive technical baselines, for example, on critical and emerging technology and supply chain security as part of Korea-US-Australia emerging-technology axis or the AUKUS Pillar II adjacency under discussion with Korea since the 2024 consultations.[52] In particular, Pillar II's focus on AI, autonomy, cyber, quantum, undersea capability directly addresses the subjects of the proposed assurance regime. So far, however, Pillar II lacks the verification layer, a crucial element that Korea could contribute (Table 1).

49 "S. Korea Joins US-Led Multinational Cyber Defense Drill"; "Cyber Operations Command Joins US-Led Cyber Flag."
50 "Trilateral Freedom Edge Exercise Wraps Off South Korea"; "US, South Korea and Japan Open Joint Air and Naval Exercise"; "PMTEC Pioneers AI-Enabled Warfare Capabilities."
51 Steff, *Small Force, Big Data*, pp. 21, 47.
52 Joint Leaders' Statement to Mark the Third Anniversary of AUKUS; Joint Statement of the 2024 United States–Republic of Korea Foreign and Defense Ministerial Meeting (2+2); 56th Security Consultative Meeting communiqué.

**Table 1: From Cross-Country Cybersecurity Information Exchange to a Defense AI Assurance Regime**

| Dimension | Existing cyber precedent | Proposed assurance regime |
|---|---|---|
| **Anchors** | Korea–US–Australia emerging-technology axis; Five Eyes adjacency | The same high-trust circle as the first ring; outward expansion as the baseline accumulates |
| **Subjects** | Time-sensitive cyber threat intelligence; joint advisories on state-backed campaigns | Evaluation results, assurance methodologies, red-team findings, failure cases |
| **Trust instruments** | Standing institution-to-institution pipelines; allied cyber exercises (Cyber Flag: ROK annually since 2022; Australia and Five Eyes at the core) | Standing assurance exchange mechanism; exercises as assurance testbeds |
| **Standards instruments** | Critical/emerging-technology and supply-chain standards consultations; AUKUS Pillar II adjacency (ROK participation discussed since 2024) | Common assurance baseline, jointly drafted in the NIST grammar; Korea contributes K-RMF integration experience |
| **Certification capacity** | Singapore's AI Verify ecosystem; commercial third-party assurance (e.g., Resaro) | Accredited third-party defense AI certification producing deployment evidence |
| **Design approach** | ASEAN cyber cooperation illustrating that voluntary, non-binding, and capacity-centered information exchange works | Non-binding baseline plus mutual recognition; capacity building for members with underdeveloped local capacities |

Source: Author

Building on this existing basis rather than from nothing carries three design consequences. First, the membership question largely answers itself at the start. A regime built on the existing cyber trust pipelines starts with Korea, the United States, Australia, and Japan, with the UK and other Five Eyes members adjacent, and Singapore, the region's certification pioneer as a natural early partner. The starting point is a minilateral first ring rather than a regional treaty. Minilateralism became the Asia-Pacific's working mode of security cooperation precisely because it trades universality for trust density, exactly what an assurance regime needs above all. The ring can widen, as the next chapter will discuss, but it does not wait for twenty countries' consensus to begin among five.

Why not simply house the regime inside an existing minilateral framework such as AUKUS Pillar II or the Quad? Membership is the problem with both options. AUKUS excludes the two allies whose sovereign AI programs make the assurance problem most acute, Korea and Japan, and its participation consultations subordinate partner involvement to a capability-cooperation agenda that treats assurance as an input, via the technology developed by the core states, rather than a jointly designed product. The Quad does not include Korea, and India's defense-AI cooperation, while long-standing with both Korea and Japan, follows a different alignment profile than the dense evidence-sharing the first ring assumes from the start. This describes who is in the first ring at the outset, not a judgment about India's longer-term role. The lesson is not that these bodies are irrelevant. The regime must work with both, and AUKUS Pillar II is its most natural standard counterpart. But for the start the first ring must be defined by a single criterion: who already exchanges technical evidence of comparable sensitivity today. No existing body matches the membership (Korea, the US, Australia, Japan, with Singapore adjacent) that criterion produces.

Second, exchange can begin at the classification level of existing information sharing mechanisms. A standard objection to allied AI assurance is that evaluation data is too sensitive to share as it reveals capabilities, vulnerabilities, doctrine. But assurance data is not threat intelligence. Most of what is exchanged today describes a common adversary or a failure mode of currently operating systems. This information already reveals insights as do the Cyber Flag exercises that expose each other's defensive weaknesses. Thus, the countries involved in current exchanges have already crossed a certain trust threshold by sharing this kind of information. However, the defense AI assurance regime discussed here, would require far less than full disclosure as assessment results, not assessment models, would be exchanged. What is needed is the extension of existing information-sharing arrangements to a new data category, not the invention of a new trust architecture.

Third, the exercise path lets the regime start on operational rather than administrative ground. The highest-leverage near-term step is an assurance annex to existing exercises, a protocol under which AI-enabled systems participating in an iteration of Freedom Edge, Cyber Flag, or Talisman Sabre arrive accompanied by declared evaluation results. For this purpose, the exercise itself would be designed to generate comparative assurance data, for example, how each nation's systems responded to common injects, where did the models diverge, and when and where did human operators override machine recommendations and why. New Zealand's experience inside Convergence and Overmatch shows partners already absorb allied AI through exercises. The proposed annex simply makes that learn-

ing explicit, structured, and shared.[53] Exercises turn assurance from a compliance activity into an operational one, and they give commanders as well as certifiers a stake in the regime.

As helpful as this approach may be, it does not supply the regime's normative design: how binding, how governed, how to expand beyond the first ring. For that, the region has its own precedent, as the next section will discuss.

53 Steff, *Small Force, Big Data*, pp. 21, 47.

# 6 The ASEAN Analogy and the Middle-Power Wager

The ASEAN Cybersecurity Cooperation Strategy 2021–2025, the second phase of a framework running since 2017, coordinates cyber cooperation among ten states of radically different regime types, threat perceptions, and capabilities. It is a rules-based multilateral order built through voluntary, non-binding norms, confidence-building measures, and coordinated capacity building.[54] Beneath the soft language the strategy commits members to exchanging information on new and existing threats, and the ASEAN Regional CERT gives that commitment standing operational form.[55] The strategy also refers to existing national CERT capability gaps and legal barriers to cross-border coordination; it is thus candid about its limits.[56] While the ASEAN Cybersecurity Cooperation Strategy does not yet address defense AI, it shows that a voluntary regional framework can help advance the exchange of critical information and gradually build trust among the partners involved. In addition, the Pacific Forum's Cyber ASEAN initiative, a regionally self-designed cyber capacity assessment framework supported by the Australian Department of Foreign Affairs and Trade built on principles of local ownership and agency, demonstrates that this region's institutions can design assessment frameworks rather than import them.[57]

One important insight follows from this approach. Dedicated capacity building was instrumental in expanding the regime. The ASEAN framework treats member capability gaps as something to be worked on through programs to close them. So should an assurance regime. The first ring builds the baseline, and a structured capacity-building track including evaluation training, shared tooling, seconded assessors widens the ring. The Philippines, for example, is the state a regional assurance regime should make a priority capacity-building partner. The analogy to ASEAN's Cybersecurity Cooperation Strategy also highlights that middle powers need a regime such as the AI assurance framework most and thus may be willing to help create it.

A superpower can self-certify: its internal evaluator market, deployment scale, and leverage over vendors make external recognition a convenience. Middle powers in the Asia-Pacific region can do none of this at scale. The European Union can project its standards abroad through the size of its market, an outcome often called the Brussels effect, but Korea has no comparable leverage. It cannot force the United States to accept regionally issued certificates through market power, so it can only make them worth accepting through evidence quality and institutional design. But middle powers have agency as the Common Criteria Recognition Arrangement shows. This framework provides a basis for mutually recognizing IT security evaluation certificates accredited by one member of the arrangement and

54 ASEAN, *Cybersecurity Cooperation Strategy 2021–2025*.
55 Ibid., para. 30 and the ASEAN Regional CERT provisions.
56 Ibid., on member capacity gaps and cross-border legal barriers.
57 Manantan (ed.), *Cyber ASEAN* (2024).

up to defined assurance levels. This solution is operated by the US in cooperation with Korea, Australia, Japan, and Singapore, with some of them certificate-authorizing members running their own accredited testing laboratories.[58] Along similar lines, the Financial Action Task Force turned a soft standards-and-peer review design into a de facto global regime for financial integrity without treaty coercion. And where the hegemon declined to lead (the Ottawa and Oslo processes on landmines and cluster munitions), middle-power coalitions built working regimes the hegemon later had to navigate. Defense AI assurance is a harder case than any of these as it touches upon more sensitive issues. But the institutional form proposed here, a voluntary baseline, accredited evaluation, and certificate recognition among trust-selected members, requires little invention. It applies the Common Criteria's form to a new evaluation object, and largely among the same states.

The wager can fail. Washington may conclude that allied assurance is a tool of dependency management, not a public good. Certificates could then become one more instrument of conditional access, as the models themselves did after the Fable 5/Mythos 5 episode. If that happens, a regional regime becomes more necessary as a hedge. But it also becomes harder to build against American ambivalence. No prediction is offered here about which way the branch goes. On either branch, however, the middle powers' dominant strategy is the same: build the recognition architecture now, anchor it in a grammar the United States already speaks, and make participation cheaper than neglect. That is the work of the next section.

58 Common Criteria Recognition Arrangement, membership and certificate-authorizing-scheme documentation.

# 7 A Working Baseline and the Road to Mutual Recognition

This section develops a common baseline for the proposed Asia-Pacific defense AI assurance regime. It builds on Korea's experience in developing the national risk framework K-RMF by combining national requirements with frameworks provided by the NIST AI RMF. This approach also points towards the integrative role Korea can play in advancing a defense AI assurance regime in the Asia-Pacific region.

## 7.1 Why This Baseline

The US federal RMF family and its DoD variant operate at scale, and K-RMF, adapted from the same NIST family to Korean defense conditions, has been adopted across the ROK military, with application proceeding in phases. This account of K-RMF draws on the author's direct professional involvement in leading its adoption; the implementing documents and the 2024 Defense AI Security Guidelines are internal ROK military documents that are not publicly released, so the account is not independently verifiable from open sources. These frameworks have already proven their value in daily use and can thus serve as a valuable foundation vetted by practitioners. In addition, the NIST family provides something like a "common language." America's own AI assurance framework is anchored in NIST.[59] Australia's assurance work and Singapore's AI Verify draw on the same measurement-and-conformity vocabulary. A baseline leveraging this foundation is readily accessible to every first-ring partner and helps create methodological interoperability.

The baseline's substantive core is the tension between security and performance that shapes operational AI. Research on adversarial machine learning has shown that robustness and accuracy can pull against each other in a range of evaluation settings.[60] How far that tension carries into operational security is mission- and threat-model dependent, which is exactly why it must be measured rather than assumed. That is why the framework needs to turn principled discussions about missions, risk tolerance, and meaningful human control into measurable metrics. What could such metrics look like? For a targeting model, they might include how far identification accuracy drops in conditions outside the training data such as fog or night, how robust the model remains when an adversary uses camouflage or spoofing to deceive it, how reliability shifts as sensor quality degrades, and what error margin a given mission type allows. None of these is a single pass-or-fail stamp, as each is a number that a partner can read and compare.

International AI evaluation standards offer a starting point for how such criteria can be structured. ISO/IEC 24029, on the robustness of neural networks, provides

59 US Department of Defense, *AI Cybersecurity RM Tailoring Guide*; NIST, *Cyber AI Profile* (2025); NIST, *AI RMF 1.0* (2023).
60 Tsipras et al., "Robustness May Be at Odds with Accuracy."

methods for testing whether performance holds under perturbation, noise, and adversarial conditions.[61] ISO/IEC TS 4213 sets out methodologies for measuring the classification performance of machine-learning models.[62] This allows allies to set the trust baseline for a given combined mission as an agreement about evidence: this mission tolerates this measured degradation envelope; that one does not. The balancing decision remains a human and political responsibility. Without it, meaningful human control over AI may only cover parts, and not the integral allied kill web, thus potentially undermining interoperability and credibility.

Finally, the baseline encodes a workable division of labor by separating what must be common from what must remain sovereign. Common work should encompass the assurance categories (robustness, degradation behavior, adversarial resistance, human-override performance), the evidence formats (what a scorecard contains, how ground-truth premises are declared), and the process requirements (independent evaluation, red-teaming, re-verification upon updates). By contrast, each nation will need to decide on its own how to define the ground truth in terms of lawfulness, doctrine, proportionality, and deployment thresholds, as discussed in section 3. Overall, partners are not asked to certify one another's values, but how they intend to work with these values within the defense AI assurance framework.

## 7.2 Roadmap

Extending the baseline into a fully-fledged regional defense AI assurance framework could be done in several steps:

- **Phase 1: Assurance Annexes (12–24 months)**
  Building on the insights of section 5, Korea should propose assurance annexes within existing exercise frameworks, such as Cyber Flag on the cyber side, Freedom Edge and successor trilateral iterations on the operational side. Participating AI-enabled systems arrive with declared evaluation scorecards in a common format. Then exercise injects would be designed to generate comparative assurance data, and after-action reviews could include an assurance line, that is, the per-system trust score drawn from the scorecard. While the costs of this approach are minimal as no new institutions need to be created, its output would be most valuable as evaluated scores for defense AI trustworthiness would, for the first time, provide a basis for cross-national comparison.

61 ISO/IEC TR 24029-1:2021, Artificial Intelligence (AI) — Assessment of the Robustness of Neural Networks — Part 1: Overview; ISO/IEC 24029-2:2023, Part 2: Methodology for the Use of Formal Methods (Geneva: ISO/IEC, 2021–2023).
62 ISO/IEC TS 4213:2022, Information Technology — Artificial Intelligence — Assessment of Machine Learning Classification Performance (Geneva: ISO/IEC, 2022).

- **Phase 2: A Standing Assurance Exchange Mechanism (24–48 months)**
  To provide a home for annex data, first ring nations would establish a standing mechanism, modeled on the cyber intelligence pipelines rather than on a treaty secretariat, to continuously exchange evaluation results, red-team findings, failure cases, and methodology updates, using unclassified extracts where raw data is too sensitive. The same approach could be used to collect ground-truth transparency by members registering their core assumptions. This provides the foundation against which every scorecard can be read and interpreted.

  Phase 2 also institutionalizes and ensures re-verification. Defense AI cannot be certified once and left alone. Models are retrained, fine-tuned, and patched on vendor and national schedules, and every update can invalidate some part of the existing evidence. The exchange mechanism allows members to verify assurance artifacts against model versions along their lifecycle, and members commit to notifying partners of relevant changes. The Fable 5 and Mythos 5 episode illustrated the same need from the opposite direction, since allied dependence on external update streams means a partner's system can change, or even vanish, without the coalition's knowledge.[63]

- **Phase 3: Accredited Certification and Mutual Recognition (48 months on)**
  With a stable baseline and a habit of evidence exchange, the third-party layer can be accredited. Two kinds of capacity can support this. The first is the independent evaluator that already tests others' systems, such as Resaro or AI Verify. It is named first because it is built for third-party certification from the outset, which is what mutual recognition requires. The second is the defense research and testing ecosystem, such as Australia's TAS and DAIRNet, which offers deep domain capacity but sits inside national defense programs. Both would have to be extended to defense grade under appropriate classification regimes, an extension this study proposes rather than one that already exists. On this basis, members could set up a mutual recognition arrangement, under which a certificate issued by one member's accredited evaluator counts as deployment evidence in the others.[64]

  Recognition can be scoped by risk tier. For example, AI for logistics and sustainment could be recognized first, followed by AI for ISR and decision support. By contrast, the assurance of AI for targeting would only be recognized as evidence that would at no point substitute for national human-control approval. This would allow nations to share evidence while keeping final judgment as a sovereign task. The tiering follows the logic developed in section 3.

63 Steff, Small Force, Big Data, pp. 7–8, 49.
64 On Resaro and AI Verify, see section 3; on Australia's TAS and DAIRNet, Layton, *Evolution not Revolution*, pp. 17, 21.

Properties that are institution-invariant, such as robustness, security behavior, degradation envelopes, can carry full certificate recognition, while functions resting on divergent national ground truths can carry evidence-grade recognition at most.

## 7.3 What Seoul Should Do

Building on its own experience, Seoul can push forward work on the proposed baseline by, first, circulating lessons from the K-RMF+AI RMF integration as a discussion paper to stimulate debate. Second, Korea should propose the first assurance annex for the next Freedom Edge iteration. Third, Seoul could designate an assurance lead with a regional-interface mandate inside the governance structure the pending defense AI legislation will create. Fourth, opening a certification-capacity dialogue with Singapore would structurally tie in the one regional partner whose conformity-assessment ecosystem is ahead of its defense applications, and strengthen the momentum for pan-regional cooperation. Taken together, the moves position Korea as a credible convener among regional middle powers.

# 8 Conclusion: Verified, Not Just Fast

There is an inherent tension between planning for allied wars and building sovereign machines. Although this tension is challenging to solve, this study argues that a regional regime advancing mutual verification of defense AI solutions can help produce evidence about what one another's machines will do and respect national caveats. The respective regime does not need to be built from scratch. Its starting point exists in the minilateral cyber trust pipelines the region's US allies already operate and the broader information exchange mechanisms ASEAN countries have established for cybersecurity. By building on these elements and filling the gaps that currently exist, middle powers in the Asia-Pacific region can demonstrate they are able to build the trust infrastructure of the algorithmic age without waiting for a hegemon to supply it.

To this purpose, countries in the region should work along six lines of effort. First, adopt a common assurance baseline that is non-binding and reflects the methodology of comparable verification frameworks such as NIST. Countries constituting the first ring such as Korea, the United States, Australia, and Japan, with Singapore as an early partner, could adopt a shared assurance baseline specifying common evidence categories, scorecard formats, and process requirements. To support these activities Korea could contribute its K-RMF+AI RMF integration lessons as one working input to a jointly drafted text. Sovereign ground truth and national deployment thresholds remain explicitly outside the baseline's scope.

Second, attach assurance annexes to existing allied exercises. Beginning with the next feasible Cyber Flag and Freedom Edge iterations, AI-enabled systems participate with declared evaluation scorecards, exercises are instrumented to generate comparative assurance data, and after-action reviews include an assurance line. Exercises are where the regime begins to operate.

Third, establish a minilateral assurance exchange mechanism. Modeled on the cyber intelligence pipelines rather than a treaty body, it exchanges evaluation results, red-team findings, and failure cases. This also maintains the baseline as a living document; and operates a ground-truth registry so that every scorecard is read against its declared evaluation standard.

Fourth, build the third-party layer by accrediting independent defense AI evaluators using the region's existing certification capacity. Mutual recognition should be built using different risk tiers. For example, verification scorecards for defense AI solutions in logistics and sustainment could pave the ground. More demanding solutions would follow using stronger verification criteria. Special categories might be needed for specific solutions. Foreign certificates confirming verification of AI for targeting, for example, could count as evidence but would never replace national judgments based on individual risk appetite.

Fifth, invest in the region's own science by jointly funding research to build indigenous evaluation knowledge that is indispensable to identify local risk tolerance, establish appropriate benchmarks, and develop ground-truth methodologies. Creating this body of expertise also gives the region's researchers an active, contributing role in global evaluation science.

Sixth, run capacity-building programs from day one. Following ASEAN's example, treat partners' lack of capacities as a program to advance evaluation training, share tools, and establish pools of seasoned and seconded assessors. Capacity-building creates an incentive and may motivate countries to join the regime rather than staying outside.

A regional verification and assurance regime built on these six lines of effort reflects the fact that military operators can entrust a system with matters of consequence only if they can explain how it fails, who can update it, and what assumptions it carries.[65] But in the Asia-Pacific region the framework for building verified trust is only slowly emerging. As this paper shows, the region's middle powers can collectively move forward building such a framework. "Moving fast and breaking things" has become a popular mantra, but it is misleading when armed forces do not understand how allied AI systems work and can undermine allied readiness to cooperate. That is why this study proposes a regional mechanism to verify allied defense AI.

One caution is needed here. Allies will not fight only with the AI they own. Most will fight with American AI, because their dependence on US defense solutions makes it the available choice. The goal is therefore not to trade owned systems for trusted ones, as that would leave most allies with nothing to field. The goal is narrower and harder: that a system an ally cannot verify is treated as an exposure rather than as trusted capability. A regional regime does not free allies from American AI. It allows them to use it, and one another's systems, based on evidence rather than assumption. This is the correction the study proposes, namely verified adoption rather than merely fast or sovereign adoption. It should also be clear what this regime is not. It does not align the region behind one great power's standard, and it does not target any single adversary. It is an instrument for allies who have already chosen to fight together, so that they can trust one another's machines. Korea's role in it is to convene, rather than to follow Washington or Brussels.

65 Steff, *Small Force, Big Data*, p. 49.

# Literature


"AI Model Rivalry Expands into Defense Sector," *The Korea Times*, 12 June 2026, https://www.koreatimes.co.kr/business/tech-science/20260612/ai-model-rivalry-expands-into-defense-sector (last accessed 14 August 2026).

"Bipartisan Military AI Bill Proposed at National Assembly," *The Korea Herald*, 2 February 2026, https://www.koreaherald.com/article/10667740 (last accessed 14 August 2026).

"Cyber Operations Command Joins US-Led Cyber Flag," *The Chosun Daily*, 27 July 2026, https://biz.chosun.com/en/en-policy/2026/07/27/5YELPIZR6RHUXLJ4U2W5OGWDPE/ (last accessed 14 August 2026).

"OpenAI and SoftBank Group partner to develop and market advanced enterprise AI," SoftBank Group, 3 February 2025, https://group.softbank/en/news/press/20250203_0 (last accessed 14 August 2026).

"PMTEC Pioneers AI-Enabled Warfare Capabilities," *Indo-Pacific Defense Forum*, 8 January 2026, https://ipdefenseforum.com/2026/01/pmtec-pioneers-ai-enabled-warfare-capabilities/ (last accessed 14 August 2026).

"S. Korea Joins US-Led Multinational Cyber Defense Drill," *The Korea Herald*, 21 July 2025, https://www.koreaherald.com/article/10536160 (last accessed 14 August 2026).

"S. Korea, UNDP Look to Next Phase of Anti-Corruption Work as AI Risks Rise," *The Korea Herald*, 7 July 2026, https://www.koreaherald.com/article/10800994 (last accessed 14 August 2026).

"South Korea's Naver, KAI to Develop Defense AI Platform," UPI, 7 July 2026, https://www.upi.com/Top_News/World-News/2026/07/07/naver-korea-aerospace-industries-defense-ai/7991783467528/ (last accessed 14 August 2026).

"Trilateral Freedom Edge Exercise Wraps Off South Korea," *USNI News*, 19 September 2025, https://news.usni.org/2025/09/19/trilateral-freedom-edge-exercise-wraps-off-south-korea (last accessed 14 August 2026).

"US, South Korea and Japan Open Joint Air and Naval Exercise," AP, 15 September 2025, https://apnews.com/article/south-korea-us-japan-freedom-edge-drills-6577be92c68446e02c10e67211db1aff (last accessed 14 August 2026).

Anthropic, "Statement on the US Government Directive to Suspend Access to Fable 5 and Mythos 5," 12 June 2026, https://www.anthropic.com/news/fable-mythos-access (last accessed 14 August 2026).

Antoniuk, Daryna, "Ukraine Says Russia Using AI Malware on Battlefield," *The Record*, 20 May 2026, https://therecord.media/ukraine-says-russia-using-ai-malware-on-battlefield (last accessed 14 August 2026).

ASEAN, *ASEAN Cybersecurity Cooperation Strategy 2021–2025* (Jakarta: ASEAN Secretariat, 2022), https://asean.org/wp-content/uploads/2026/07/ASEAN-Cybersecurity-Cooperation-Strategy_08072026_FIN.pdf (last accessed 14 August 2026).

Bimo, Emanuel Ario, and Bagus Jatmiko, *A Careful Walk on Thin Ice: Defense AI in Indonesia*. DAIO Study 25|31 (Hamburg: Defense AI Observatory, 2025), https://defenseai.eu/daio_study2531_a_careful_walk_on_thin_ice_emanuel_ario_bimo_bagus_jatmiko (last accessed 14 August 2026).

Borchert, Heiko, "Ambivalent Acceleration: War, Battlefield Deployment, and Money Push Defense AI But Fundamental Questions Remain Unresolved" (18 June 2026), forthcoming in Frank Schmiedchen et al. (eds.), *Living with AI: State, Applications, and Challenges of Artificial Intelligence* (Berlin: Springer); preprint at SSRN, https://ssrn.com/abstract=7122758 (last accessed 14 August 2026).

Borchert, Heiko, "The Very Long Game of Defense AI Adoption: Introduction," in Heiko Borchert, Torben Schuetz, and Joseph Verbovszky (eds.), *The Very Long Game: 25 Case Studies on the Global State of Defense AI* (Cham: Springer, 2024), pp. 1-38, https://link.springer.com/book/10.1007/978-3-031-58649-1 (last accessed 14 August 2026).

Chavez, Pablo, Ruby Scanlon, and Vivek Chilukuri, "Sovereign AI Index" (Washington, DC: Center for a New

American Security, 21 April 2026), https://interactives.cnas.org/reports/sovereign-ai-index/ (last accessed 14 August 2026).

Chiba, Kanako and Ryo Nemoto, "Japan weighs use of Palantir, Anduril AI systems for defense forces," Nikkei Asia, 5 August 2026, https://asia.nikkei.com/politics/defense/japan-weighs-use-of-palantir-anduril-ai-systems-for-defense-forces (last accessed 14 August 2026).

Common Criteria Recognition Arrangement, https://www.commoncriteriaportal.org (last accessed 14 August 2026).

DARPA, "AI Forge: Building the Assurance Foundations of National Security AI," program materials, June 2026, https://www.darpa.mil/research/programs/ai-forge (last accessed 14 August 2026).

European Union, Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act), *Official Journal of the European Union*, 13 June 2024, https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (last accessed 14 August 2026).

Infocomm Media Development Authority (Singapore), AI Verify framework documentation (last accessed 14 August 2026).

Jensen, Benjamin, Ian Reynolds, Yasir Atalan, Michael Garcia, Austin Woo, Tony Chen, and Trevor Howarth, "Critical Foreign Policy Decision (CFPD) Benchmark: Measuring Diplomatic Preferences of Large Language Models" (Washington, DC: Center for Strategic and International Studies and Scale AI, 2026), https://www.csis.org/analysis/asking-ai-how-respond-foreign-policy-decision (last accessed 14 August 2026).

Joint Leaders' Statement to Mark the Third Anniversary of AUKUS (17 September 2024), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2024/09/17/joint-leaders-statement-to-mark-the-third-anniversary-of-aukus/ (last accessed 14 August 2026).

Joint Statement of the 2024 United States–Republic of Korea Foreign and Defense Ministerial Meeting (2+2) (Washington, DC: US Department of State, 31 October 2024), https://2021-2025.state.gov/joint-statement-of-the-2024-united-states-republic-of-korea-foreign-and-defense-ministerial-meeting-22/ (last accessed 14 August 2026).

Jung, Ryu, "Japan's Self-Defense Forces Adopt Palantir AI for Command," The Chosun Daily, 29 June 2026, https://www.chosun.com/english/world-en/2026/06/27/HLUOX54EBRAZ5PEFX7DF2TUAMA/ (last accessed 14 August 2026).

Layton, Peter, *Evolution not Revolution: Australia's Defence AI Pathway*. DAIO Study 22|02 (Hamburg: Defense AI Observatory, 2022), https://defenseai.eu/daio_study2202 (last accessed 14 August 2026).

Manantan, Mark Bryan F. (ed.), *Cyber ASEAN: Advancing Cyber Resiliency and Capacity in Southeast Asia* (Honolulu: Pacific Forum, March 2024), https://pacforum.org/wp-content/uploads/2024/03/CyberAsean_Report.pdf (last accessed 14 August 2026).

Microsoft, "Microsoft Deepens Its Commitment to Japan with $10 Billion Investment in AI Infrastructure, Cybersecurity, and Workforce," 3 April 2026, https://news.microsoft.com/source/asia/2026/04/03/microsoft-deepens-its-commitment-to-japan-with-10-billion-investment-in-ai-infrastructure-cybersecurity-workforce/ (last accessed 14 August 2026).

NATO, "Summary of the Artificial Intelligence Strategy," 22 October 2021, https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2021/10/22/summary-of-the-nato-artificial-intelligence-strategy (last accessed 14 August 2026).

NIST, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (Gaithersburg, MD: National Institute of Standards and Technology, 2023), https://www.nist.gov/itl/ai-risk-management-framework (last accessed 14 August 2026).

NIST, *Cyber AI Profile*, NIST IR 8596 (initial public draft) (Gaithersburg, MD: National Institute of Standards and Technology, December 2025).

Park, Youngwook, *Will the One Ring Hold? Defense AI in South Korea*. DAIO Study 24|21 (Hamburg: Defense AI Observatory, 2024), https://defenseai.eu/daio_study2421_will_the_one_ring_hold_youngwook_park (last accessed 14 August 2026).

Resaro, "Resaro Levels Up: Establishing Co-Headquarters in Europe," January 2025, https://resaro.ai/insights/resaro-establishes-co-headquarters-in-europe (last accessed 14 August 2026).

Steff, Reuben, *Small Force, Big Data: Defense AI in New Zealand*. DAIO Study 26|35 (Hamburg: Defense AI Observatory, 2026), https://defenseai.eu/daio_study2635_small_force_big_data_reuben_steff (last accessed 14 August 2026).

Sygaco, Keanu Paul B., Reimagining Horizons. Defense AI in the Philippines (Hamburg: Defense AI Observatory, forthcoming).

The White House, National Security Presidential Memorandum/NSPM-11, "Artificial Intelligence in the National Security Enterprise" (Washington, DC, 5 June 2026), https://www.whitehouse.gov/presidential-actions/2026/06/national-security-presidential-memorandum-nspm-11/ (last accessed 14 August 2026).

Tsipras, Dimitris et al. "Robustness May Be at Odds with Accuracy," International Conference on Learning Representations (2019), arXiv:1805.12152, 9 September 2019, https://arxiv.org/abs/1805.12152 (last accessed 14 August 2026).

Tsuchiya, Motohiro, *Overcoming the Long Shadow of the Past: Defense AI in Japan*. DAIO Study 24|23 (Hamburg: Defense AI Observatory, 2024), https://www.defenseai.eu/daio_study2423_overcoming_the_long_shadow_of_the_past_motohiro_tsuchiya (last accessed 14 August 2026).

UK Department for Science, Innovation and Technology, statement on the renaming of the AI Safety Institute as the AI Security Institute (Munich Security Conference, 14 February 2025).

UK National Cyber Security Centre, "Why Cyber Defenders Need to Be Ready for Frontier AI," NCSC blog, 30 March 2026, https://www.ncsc.gov.uk/blogs/why-cyber-defenders-need-to-be-ready-for-frontier-ai (last accessed 14 August 2026):

US Department of Defense, *Artificial Intelligence Cybersecurity Risk Management Tailoring Guide*, Version 2 (Washington, DC: Department of Defense, 2025), https://www.dmi-ida.org/download-pdf/pdf/AI-CybersecurityRMTailoringGuide.pdf (last accessed 14 August 2026).

Wang, Wenxuan, et al., "All Languages Matter: On the Multilingual Safety of Large Language Models," arXiv:2310.00905, 20 June 2024, https://arxiv.org/abs/2310.00905 (last accessed 14 August 2026).

Yong, Zheng-Xin, Cristina Menghini, and Stephen H. Bach, "Low-Resource Languages Jailbreak GPT-4," arXiv:2310.02446, 27 January 2024, https://arxiv.org/abs/2310.02446 (last accessed 14 August 2026).

## Defense AI Observatory Studies

**26|36** João Vicente, Turning Experimentation into Military Advantage. Defense AI in Portugal

**26|35** Reuben Steff, Small Force, Big Data. Defense AI in New Zealand

**26|34** João Paulo Moralez and Santiago Rivas, Fragmented Efforts. Defense AI in Brasil

**26|33** Daniel Fiott, Controlled Adaptation and Integration. Defence AI in Belgium

**26|32** Raluca Csernatoni, From Laggard to Potential Regional Innovator? Defence AI in Romania

**25|31** Emanuel Ario Bimo and Bagus Jatmiko, A Careful Walk on Thin Ice. Defense AI in Indonesia

**25|30** Heiko Borchert, Bonds That Separate. Defense AI in the Arab Gulf States

**25|29** Rafał Lipka, Jakub Wernik, and Wojciech Dzięgiel, Guarding the Alliance's Eastern Frontiers. Defense AI in Poland

**25|28** Umaima Ali, An Underdeveloped Strategic Necessity. Defence AI in Pakistan

**25|27** Heiko Borchert, Christian Brandlhuber, and Jeronimo Dzaack, Thinking Before Sinking. Red and Blue AI Tactics for Naval Combat

**24|26** Heiko Borchert and Christian Brandlhuber, His Hands Can't Hit What His Eyes Can't See. Emergent AI Tactics for Successful Operations in the Air Littoral

**24|25** Mahmoud Javadi, Heavy Thunder, No Rain: Defense AI in Iran

**24|24** Kitsch Liao, Intelligent National Defense Amid Strategic Ambiguity? Defense AI in Taiwan

**24|23** Motohiro Tsuchiya, Overcoming the Long Shadow of the Past. Defense AI in Japan

**24|22** Vitaliy Goncharuk, Survival of the Smartest? Defense AI in Ukraine

**24|21** Youngwook Park, Wil the One Ring Hold? Defense AI in South Korea

**23|20** Tomas Jermalavičius, Caught Between Today and Tomorrow. Defence AI in Estonia

**23|19** Nikolaos Karampekios, Konstantinos Sakalis, and Iraklis Oikonomou, Harnessing the Potential. Defense AI in Greece

**23|18** Andreas Graae, Servers Before Tanks? Defence AI in Denmark

**23|17** Kévin Martin and Lucie Liversain, A Winding Road Before Scaling-Up? Defense AI in France

**23|16** Sami O. Järvinen, Cautious Data-Driven Evolution. Defence AI in Finland

**23|15** Inbar Dolinko and Liran Antebi, Embracing the Organized Mess. Defense AI in Israel

**23|14** Alastair Finlan, A Fertile Soil for AI? Defense AI in Sweden

**23|13** John Lee, "Overtaking on the Curve?" Defense AI in China

**23|12** Heiko Borchert, Torben Schütz, and Joseph Verbovszky, Master and Servant. Defense AI in Germany

**23|11** Katarzyna Zysk, High Hopes Amid Hard Realities. Defense AI in Russia

**23|10** Yvonne Hofstetter and Joseph Verbovzsky, How AI Learns the Bundeswehr's "Innere Führung." Value-Based Engineering with IEEE7000™-2021

**23|09** Robert C Engen, When the Teeth Eat the Tail: A Review of Canada's Defence Artificial Intelligence

**23|08** Çağlar Kurç, Enabling Technology of Future Warfare. Defense AI in Turkey

**23|07** Lauren A. Kahn, Risky Incrementalism. Defense AI in the United States

**22|06** Yvonne Hofstetter, Wie KI Innere Führung lernt. Wertbasierte Technik mit IEEE7000™-2021

**22|05** Andrea Gilli, Mauro Gilli, and Ivan Zaccagnini, Exploring the Benefits of a New Force Enabler: Defense AI in Italy

**22|04** Kenneth Payne, Bright Prospects – Big Challenges. Defense AI in the United Kingdom

**22|03** Heiko Borchert, Christian Brandlhuber, Armin Brandstetter, and Gary S. Schaal, Free Jazz on the Battlefield. How GhostPlay's AI Approach Enhances Air Defense

**22|02** Peter Layton, Evolution not Revolution. Australia's Defence AI Pathway

**21|01** Heiko Borchert, Torben Schütz, Joseph Verbovszky, Beware the Hype. What Military Conflicts in Ukraine, Syria, Libya, and Nagorno-Karabakh (Don't) Tell Us About the Future of War

Defense AI Observatory

www.defenseai.eu