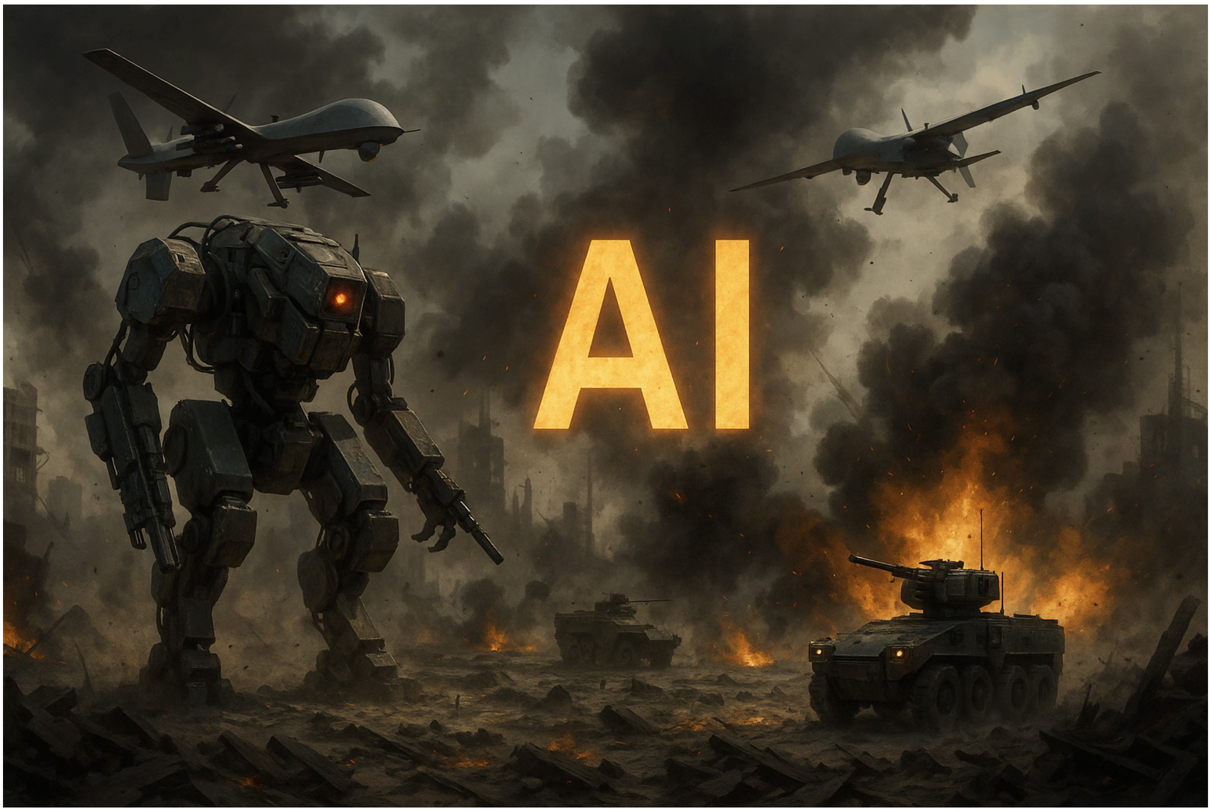
the conflict ledger

CLASSIFIED

Special report: Autonomous Systems and Artificial Intelligence in Modern Warfare



Date of Production: August 11, 2025	This briefing document synthesizes information from various sources to provide a detailed overview of the development and implications of Lethal Autonomous Weapon Systems (LAWS), and the critical ethical and security challenges posed by AI in the military domain.

Subject Details

Overview of the evolving landscape of military technology:

The modern battlefield is undergoing a rapid transformation driven by advanced weapon systems and the increasing integration of Artificial Intelligence (AI).

Active Protection Systems (APS): These are a component of modern armored vehicle defense, designed to intercept and neutralize incoming threats before they impact the vehicle. These "hard-kill" systems typically involve sensors to detect incoming projectiles and countermeasures to destroy or disrupt them.

Key APS Technologies

1. **AMAP-ADS / StrikeShield (Germany)**

Source: conflict ledger via stock images

Developer: ADS Gesellschaft für aktive Schutzsysteme, a joint venture between Rheinmetall and IBD Deisenroth Engineering.

Design & Operation: A modular, hard-kill APS with sensor-countermeasure modules arranged around the vehicle. A processor determines the threat's trajectory, activating a countermeasure near the predicted impact point to destroy or disrupt the threat. The system boasts a short reaction time of approximately 560 microseconds, eliminating threats at ranges of about 10 meters. It minimizes collateral damage through the use of "non-fragmenting stream of material."

Weight: 140 kg for light vehicles, up to 500 kg for heavy vehicles.

Effectiveness: Proven against RPG-7s and ATGMs in various demonstrations, including urban combat scenarios, with "zero residual penetration." It can detect a 7.62mm round but "rejected it as a threat."

Variants:ADS-Gen3 (2017): Offers improved low-power radar sensors (20-30 GHz waveband, 1 Watt output), reducing EM signature and allowing more accurate threat determination (e.g., against RPG-30). Certified to the highest safety standard (IEC 61508) in 2017. Interception is claimed to be "so accurate that the warhead of an incoming projectile can be defeated without setting off the fuse, resulting in less danger to nearby civilians, infantry and vehicles."

StrikeShield (2019): A "hybrid protection module" integrating APS components into passive spaced armor, marketed internationally. It can be mounted on existing passive armor interfaces, saving up to 35% of overall system weight.

Countermeasures: Both ADS-Gen3 and StrikeShield can use a lighter effector for ATGMs and RPGs, and a heavier effector for armor-piercing rounds like large-calibre APFSDS ammunition. Tests showed StrikeShield could fragment or tilt APFSDS projectiles, potentially reducing penetration by "up to 75%."

Applications: Tested on Marder, SEP, Combat Vehicle 90, Patria AMV, Iveco LMV. In series production for Hungarian KF41 Lynx and used on Singapore Army Leopard 2 tanks (over 80 systems delivered). Showcased on KF51 Panther and GTK Boxer.

Operational advantage: Faster than Quick Kill, Iron Fist, or Trophy. No moving parts, reducing weight and power requirements.

2. Arena (Russia)

Source: Vitaly V. Kuzmin

Developer: Kolomna Engineering Design Bureau (KBM), Russia.

Purpose: Protects armored fighting vehicles from light anti-tank weapons, ATGMs, and flyover top attack missiles.

Operation: Uses a Doppler radar to detect incoming warheads, then fires a defensive rocket that detonates "near the inbound threat, destroying it before it hits the vehicle."

Mass: 1100 kg.

Background: Developed in the early to mid-1990s following heavy Russian casualties from RPGs in Chechnya. It evolved from the earlier Drozd (1970s) and Shtora (1980s) systems. Arena is a "hard-kill system like Drozd," unlike Shtora which was a "soft-kill system" designed to jam guidance systems.

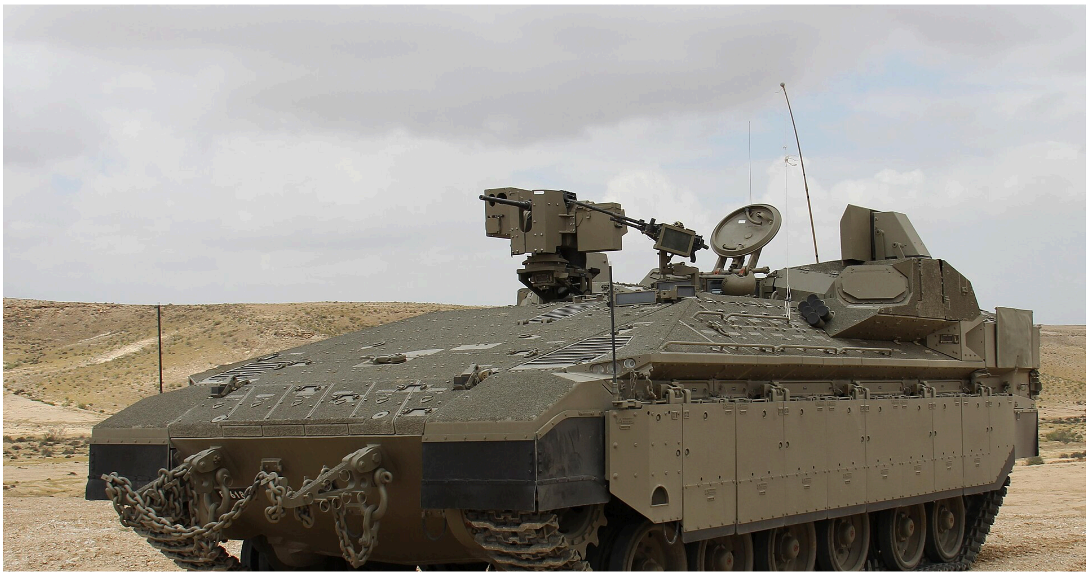
System Details: Uses a multi-function Doppler radar. A digital computer selects one of 26 quick-action projectiles to intercept the threat within 0.05 seconds. Protects over a 300-degree arc on the T-80UM (excluding the rear) and 260 degrees on the T-72M1. Engages targets within 50 meters, with ammunition detonating around 1.5 meters from the threat. Can engage threats traveling between 70 m/s and 700 m/s, disregarding false targets like birds or small caliber bullets.

Effectiveness: Increases tank survival probability against RPGs by "between 1.5–2 times."

Variants: Arena-M: Modernized version claimed to intercept munitions from all aspects, including top-attack missiles like the Javelin. Slated for installation on T-80 and T-90 tanks, with exploration for T-72B3 and T-72B3M.

Exports: General Dynamics Land Systems (GDLS) proposed licensing Arena for M60-2000 and M1A2 Abrams in 1998. South Korea agreed to fit Arena-E on K2 main battle tank in 2007 (US$27.5 million).

3. **Trophy (Israel)**



A Namer AFV equipped with Rafael's Trophy system.

Developer: Rafael Advanced Defense Systems (Israel).

Purpose: Active protection system for military armored vehicles, supplementing standard armor and enhancing enemy location identification.

Design & Operation: Centered on the Elta EL/M-2133 F/G band fire-control radar with four flat-panel antennas providing a 360-degree sensing field. Upon detection, it computes threat parameters and launches small Explosively Formed Penetrators (EFPs) to form a "precise and closely spaced matrix, targeting an area in front of the anti-tank projectile." Designed with a "narrow kill zone to ensure the safety of friendly personnel."

Capabilities: Protects against ATGMs, rockets, HEAT rounds, RPGs, and recoilless rifles. Can engage numerous threats simultaneously from different directions, effective on stationary or moving platforms. Newer versions include automated reloading. Can identify if a threat will miss the platform, providing shared location data without activating countermeasures.

Mass: 820 kg (Trophy HV), 480 kg (Trophy MV/VPS).

Combat History: Declared operational by Israeli Defense Forces in August 2009. First operational on March 1, 2011, foiling a missile attack on a Merkava MK IV tank. During Operation Protective Edge (2014), it performed "over a dozen interceptions of anti-tank weapons including Kornet, Metis, and RPG-29," with "no tanks damaged."

Limitations: Adds significant weight (half a ton for core system). Currently incapable of defeating kinetic energy anti-tank weapons. Risks to dismounted infantry from EFPs. Has a "donut-hole like window of vulnerability to attacks directly from above," allowing drones to drop grenades (e.g., Hamas in October 2023). Can be overwhelmed by tactics like RPG-7s within 50m, weapons exceeding sound speed (SPG-9), or multiple rounds in quick succession.

Variants: Trophy MV/VPS (formerly Trophy Light): Designed for light and medium armored vehicles (e.g., Stryker, Bradley). Approximately 40% lighter and smaller. Achieved over 95% success rate in 2018 qualification tests.

Trophy LV: Even lighter (200 kg) for vehicles under 8 tons (e.g., jeeps, 4x4s).

International Operators: In service with Germany (Leopard 2A7 A1, Leopard 2A8), United States (Abrams M1A1/A2, tested on Stryker and Bradley), and selected by UK for Challenger 3 tanks. India's Larsen & Toubro signed MoU to manufacture Trophy in India.

Sentry Guns: Automated Point Defense

These are stationary or mobile weapon systems that automatically detect, aim, and fire at targets using sensors.

Source: Conflict ledger via stock images

A prominent example of a close-in weapon system (CIWS), operational since 1980.

Purpose: Automated gun-based system to defend military watercraft against incoming threats like aircraft, missiles, and small boats. Land variant (LPWS, part of C-RAM) counters rocket, artillery, and mortar attacks.

Design: Radar-guided 20 mm Vulcan cannon mounted on a swiveling base. Fully self-contained, capable of "automatically search for, detect, track, engage, and confirm kills."

Operation: Uses a search radar to identify targets, then a tracking antenna for precise engagement. Fires 4,500 rounds/minute of 20 mm armor-piercing tungsten penetrator rounds. Tracks outgoing rounds to "walk" them onto the target.

Autonomy: Does not recognize IFF (identification friend or foe). Decides to engage based on real-time radar data, considering factors like range, maneuverability, and velocity. The system can function despite significant ship damage, requiring minimal external inputs.

Incidents: Historically, there have been accidental firings and friendly fire incidents (e.g., USS Jarrett hitting USS Missouri due to chaff in 1991 Gulf War; Japanese destroyer hitting US aircraft in 1996).

Combat Use:

a. used by USS Gravely in January 2024 to shoot down a Houthi anti-ship missile in the Red Sea, marking its first operational downing of a Houthi missile.

b. **Samsung SGR-A1 (South Korea):** Military robot sentry for the demilitarized zone at the South and North Korean border.

c. **Sentry Tech (Israel):** Deployed along the Gaza border fence, mounts a .50 BMG automated M2 Browning machine gun and SPIKE guided missile. Operates with human input to fire but can acquire and track targets independently.

d. **Super aEgis II (South Korea):** Automated turret-based weapon platform using thermal imaging to lock onto vehicles or humans up to 3 km away. Capable of automatic firing but customers configure it for human confirmation.

e. **Bullfrog (United States):** AI-enabled sentry gun system tested in 2024, consisting of an M240 machine gun on a rotating turret with electro-optical sensor, proprietary AI, and computer vision software.

f. **Sky Sentinel (Ukraine):** AI-operated turret unveiled in May 2025, fitted to an M2 Browning machine gun, which "shot down its first Shahed drone without human input using radar." Costs around $150,000 per turret.

Lethal Autonomous Weapon Systems (LAWS) and the Role of AI

LAWS are military drones or robots that can independently search for and engage targets based on programmed constraints. While rudimentary autonomous functions (like landmines or CIWS) have existed for decades, newer systems are more sophisticated and increasingly incorporate AI.

Defining Autonomy and LAWS

No Universal Definition: There is "no commonly agreed definition of Lethal Autonomous Weapon Systems (LAWS)."

US Department of Defense: Defines an Autonomous Weapons System as "A weapon system that, once activated, can select and engage targets without further intervention by a human operator."

Heather Roff: Describes them as "armed weapons systems, capable of learning and adapting their 'functioning in response to changing circumstances in the environment in which [they are] deployed,' as well as capable of making firing decisions on their own."

Scholarly View: Scholars like Peter Asaro and Mark Gubrud consider any weapon system capable of releasing lethal force "without the operation, decision, or confirmation of a human supervisor" to be autonomous.

> Human-Control Classifications:Human-in-the-loop: Human must instigate the action (not fully autonomous).
>
> Human-on-the-loop: Human may abort an action.
>
> Human-out-of-the-loop: No human action is involved.

AI's Role in Autonomous Systems

AI is not a prerequisite for autonomous weapons, but it "could further enable such systems."

Autonomous capabilities can be provided via "pre-defined tasks or sequences of actions based on specific parameters," or through AI tools that "derive behavior from data, thus allowing the system to make independent decisions or adjust behavior based on changing circumstances."

AI can also serve in an "assistance role," for example, a computer vision system using AI to "identify and draw attention to notable objects...without having the capacity to respond to those objects autonomously."

Examples of Autonomous Offensive Systems

Loitering Munitions ("Suicide Drones"): Contain a warhead and "wait (loiter) around a predefined area until a target is located by an operator on the ground or by automated sensors onboard, and then attacks the target." Their functionalities have become "increasingly sophisticated, allowing for, among other things, longer ranges, heavier payloads and the potential incorporation of artificial intelligence (AI) technologies."

Purpose: Elevate human pilots to "mission commanders," with AIs as "loyal wingmen" for tactical control of low-cost robotic craft. They can act as "a sensor, as a shooter, as a weapons carrier, as a cost reducer."

DARPA AlphaDogfight Trials (2020): Established that AI programs piloting fighter aircraft "will overmatch human pilots." The X-62A VISTA testbed has demonstrated AI-controlled dogfighting capability.

Funding: The US Air Force plans to spend over $8.9 billion on CCA programs from FY2025-2029, with an initial goal of 1,000 CCAs at $25-30 million per airframe.

Programs: USAF Next Generation Air Dominance (NGAD) program, Skyborg manned-unmanned programs (e.g., Autonomous Air Combat Operations, AACO), DARPA Air Combat Evolution (ACE), Longshot (air-launched UAV for extended range and reduced risk to manned aircraft).

Some known suicide drones include:

a. Russia: Developing AI-enabled missiles, drones, unmanned vehicles, and military robots.
b. China: Developing "loyal wingman" prototypes (e.g., AVIC Dark Sword, Hongdu GJ-11, Chengdu WZ-10) and aims for

twin-seat stealth fighters (J-20S, J-36) to command drone swarms. China also unveiled the Feihong FH-97 prototype UCAV as a "loyal wingman" drone for various tasks, including electronic countermeasures, early warning, and decoy operations.

c. Israel: Minister Ayoob Kara stated in 2017 that Israel is developing military robots, "including ones as small as flies." In May 2021, Israel conducted an "AI guided combat drone swarm attack in Gaza."

d. Turkey: TAI Anka-3 stealth UCAV completed maiden flight in Dec 2023. In Oct 2024, it became the "first drone in history to be controlled by another aircraft in the loyal wingman role."

e. United States:Loyal Wingman / Collaborative Combat Aircraft (CCA): UCAVs incorporating AI designed to collaborate with next-generation crewed combat aircraft. Expected to be "significantly lower-cost than a crewed aircraft with similar capabilities" while increasing survivability.

The United Nations and LAWS

Secretary-General's Stance: Since 2018, António Guterres has called LAWS "politically unacceptable and morally repugnant" and advocated for their prohibition under international law by 2026. He notes that without specific multilateral regulations, LAWS "raise humanitarian, legal, security and ethical concerns and pose a direct threat to human rights and fundamental freedoms."

UN General Assembly Resolution (Dec 2023/2024): Adopted a resolution supporting international discussion on LAWS concerns, moving the debate beyond just lethal autonomous weapon systems to "the wide range of military applications of AI."

CLASSIFIED

ICRC's Concern: LAWS "pose humanitarian risks, legal challenges and ethical concerns due to the difficulties in anticipating and limiting their effects." They increase dangers for civilians and dismounted soldiers, and can "accelerate the use of force beyond human control," risking unpredictable escalation.

ICRC Recommendations:

a. Prohibit unpredictable autonomous weapons: Those whose effects "cannot be sufficiently understood, predicted and explained," including 'learning' systems and potentially all machine learning-controlled LAWS.
b. Prohibit autonomous weapons designed to apply force against people directly.
c. Strict restrictions on all other autonomous weapons to mitigate risks, ensure legal compliance, and address ethical concerns.
d. The ICRC believes a "new legally binding treaty" is needed, possibly a Protocol to the Convention on Certain Conventional Weapons (CCW).

Challenges of AI in the Military Domain

The integration of AI in military contexts presents both complex challenges across technological, security, legal, policy, and ethical dimensions.

a. **Technological Challenges**:

Data Quality and Availability: AI algorithms require "vast amounts of training data," which can be "scarce, incomplete or biased" in military contexts, leading to "unpredictable and potentially harmful outcomes." Bias in data can "reproduce or even amplify those biases in its outputs," potentially causing misidentification of targets or civilians.

Opacity ("Black Box" Nature): Many AI models are not "explainable to human operators," making it hard to assess trustworthiness, diagnose errors, and complicate accountability. This lack of transparency "erodes human confidence."

Testing and Evaluation: AI systems are adaptive, requiring "continuous evaluation" and "iterative legal reviews." Performance is context-dependent, and "non-transferability of performance" means a system effective in one environment may not be in another. Comprehensive testing is complex, especially for "systems of systems."

Cybersecurity: AI systems are vulnerable to attacks (e.g., "data poisoning," "adversarial evasion attacks," "sponge attacks," "model extraction," "model inversion," "membership inference"). Vulnerabilities could be exploited by adversaries, necessitating "robustness against spoofing or manipulation."

Misuse or Misunderstanding: Operators may "over-rely on AI recommendations" (automation bias) or "distrust and ignore AI entirely" (algorithm aversion). Poor interfaces or insufficient training exacerbate these issues.

b. Security Challenges:

Unintended Escalation and Loss of Human Agency: High-speed AI-enabled systems could escalate conflicts too rapidly for human intervention ("flash wars" or "algorithmic interactions").

AI Arms Race: "Major powers are investing heavily in military AI to avoid falling behind rivals," leading to rapid, premature deployment of unproven technologies and increased "mistrust and the likelihood of confrontation." This also encourages the use of the "battlefield as a testing ground for novel AI capabilities."

Proliferation: "Commercial off-the-shelf or open-source AI tools can be repurposed by non-state actors, terrorist groups or other armed groups," altering the threat landscape and necessitating robust "life-cycle management of military AI systems, including strict decommissioning protocols."

Information Environment Disruption: Generative AI can produce "disinformation at scale," eroding trust in information and institutions, potentially "destabilizing societies" and impacting military operations.

c. **Legal, Policy, and Ethical Challenges:**

Legal Compliance: Ensuring AI use complies with international law (IHL, human rights law, criminal law). Key debates include accountability and responsibility for AI-driven actions, particularly lethal decisions. The "accountability gap" arises when incidents occur (e.g., AI misclassifies a target), making it difficult to attribute responsibility.

Existing vs. New Legal Frameworks: Some states argue existing IHL is sufficient, while others believe AI's speed and autonomy require "new, dedicated rules."

"Meaningful Human Control": A highly debated concept, often proposed as a means to satisfy legal and ethical requirements, but its definition and application remain unsettled.

Legal Reviews (Article 36): Applying Article 36 of Additional Protocol I to AI systems is challenging, requiring review of algorithms and data, not just hardware, potentially through "iterative legal reviews."

Policy and Governance: Many countries are only beginning to draft national AI strategies for military use. Internationally, there's no dedicated intergovernmental process for military AI, leading to fragmented discussions and "governance loopholes." Achieving consensus is difficult due to varying perspectives and geopolitical contexts.

Ethical Concerns: "Widespread and serious concerns over ceding life-and-death decisions to sensors and software." This "dehumanizing process" undermines human values and moral agency. AI systems can inherit and amplify societal biases (e.g., gender, race) from training data, potentially leading to discriminatory outcomes in targeting.

Multi-Stakeholder Involvement: Much AI innovation comes from the private sector and academia, necessitating their input and cooperation in governance, but bridging the gap between national security and open technology communities is challenging.

Recommendations for AI Governance (UNIDIR):

A comprehensive roadmap is needed, with actions at multilateral, regional, and national levels:

Multilateral:

Establish a central comprehensive platform for dialogue on military AI's broader implications for peace and security.

Develop core principles for responsible AI in the military domain, drawing from existing frameworks (UNESCO, Global Digital Compact, CCW).

Further develop these principles into "international voluntary norms or guidelines for responsible state behaviour," potentially a code of conduct or political declaration.

Develop Confidence-Building Measures (CBMs) for military AI (e.g., information exchanges, notification regimes, joint expert groups, incident reporting).

Promote multi-stakeholder engagement (industry, academia, civil society) in discussions.

Implement coherent capacity-building programs to ensure inclusive rule development and responsible technology adoption.

Regional:

Leverage existing regional organizations and frameworks to discuss military AI, develop region-specific CBMs/norms, and facilitate information sharing.

Initiate cross-regional dialogues to foster mutual learning and align approaches.

National:

Formulate and implement a comprehensive national strategy on AI in security and defense, outlining vision, priorities, governance, and compliance with international law and ethics.

Establish robust governance structures and review processes (e.g., AI steering committees, ethics review boards, iterative legal reviews).

Implement transparency and accountability measures (e.g., detailed logs of AI system decisions, clear accountability protocols for commanders).

Prioritize data governance and quality, investing in high-quality, representative, and disaggregated data sets, and establishing responsible data collection and lifecycle procedures.

Adopt a life-cycle management approach for AI capabilities (design, development, testing, deployment, updates, decommissioning), including rigorous AI assurance processes.

Invest in human capital and training for military personnel on technical, ethical, and legal aspects of AI use, incorporating tailored scenarios into exercises.

Review military operational guidelines (doctrines, SOPs, TTPs, rules of engagement) to account for AI's impact and ensure clear accountability.

This report was produced by the AI Salon consortium. Lead contributors: Mwende Mukwanyaga, Okari M., and Hesbon Ombati, with pattern identification support from Notebook LM. Edited by Lilian Mutinda. Cover image by Sora.

END REPORT