OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Shifting the Balance of Stability: The Automation Gap, AI-Defense, and the Transition to First-Strike Cyber Exploitation

Ezra Matiwos Wesenie
ezramatiwos1@gmail.com

ABSTRACT

Historically, state-sponsored cyber operations operated within a 'grey zone'—defined here as an operational and diplomatic sanctuary of sub-threshold conflict, where low-intensity digital intrusions remain below the threshold of conventional armed conflict or formal military response. However, autonomous, AI-driven cyber defense platforms (such as automated, enterprise-wide AEGIS security architectures) have closed the "Automation Gap," which represents the traditional time delay between an attacker exploiting a vulnerability and a defender deploying a patch. Operating at machine speed, automated defenders instantly detect and neutralize low-intensity intrusions. This capability systematically dismantles the stabilizing temporal buffer that once protected peacetime statecraft, forcing cyber operations out of the grey zone.

Applying three core theoretical frameworks, this paper examines how defensive AI dominance paradoxically destabilizes international relations. First, it uses the **Security Dilemma 2.0**, where a state installing powerful defensive AI makes neighboring states feel threatened, driving them to prepare aggressive counter-attacks. Second, it incorporates the **Deception-Detection Dichotomy**, which highlights the structural divide between AI's high proficiency in pattern-based threat detection and its weakness in executing creative, stealthy attacks. Third, it applies the **Subversive Trilemma**, the strategic constraint holding that a cyberattack can never maximize speed, intensity, and stealth simultaneously.

Automated defense renders gradual digital subversion ineffective. Consequently, it incentivizes revisionist states—those seeking to alter the geopolitical status quo—to bypass traditional escalation ladders entirely. These states are driven to launch high-stakes "first-strike" cyberattacks that cause physical destruction (cyber-kinetic campaigns). Their goal is to achieve an unalterable victory (a *fait accompli*) before automated defenses can adapt or human diplomats can intervene.

This study demonstrates how competing strategic paradigms accelerate escalation. It compares the offensive, proactive "Persistent Engagement" model of the United States with the defensive, denial-oriented "Digital Fortress" model of Singapore. The analysis leverages case studies from Operations Epic Fury, Roaring Lion, and Cyber Guardian. Furthermore, the paper shows how these

technological gaps force small states to surrender digital control to superpower-led "AI-Defense Blocs." Meanwhile, international governance efforts—such as the UN permanent Global Mechanism launched in March 2026—remain constrained by enforcement gaps in international law (the "Legal Trilemma"). Finally, the study highlights how automated cyberattacks on dual-use conventional and nuclear command networks risk triggering preemptive nuclear escalation under "use-it-or-lose-it" pressures.

To restore strategic stability, this paper proposes three critical policy interventions: cryptographic machine-speed hotlines to exchange automated tokens of intent during high-velocity network events, mandatory human-in-the-loop agentic AI guardrails to require explicit human authorization for high-risk operations, and decentralized AI-defense capacity-building to democratize security tools for developing nations.

CHAPTER 1: INTRODUCTION

The evolution of cyberspace as a primary domain of geopolitical competition has historically been characterized by its utility as a strategic sanctuary for low-stakes, deniable conflict. For nearly three decades, state-sponsored cyber operations existed within a digital 'grey zone.' This space is defined by strategic analysts as an arena of constant, competitive interaction deliberately calibrated to remain below the threshold of conventional military response or formal declarations of war. Within this environment, states engaged in continuous espionage, intellectual property theft, and localized sabotage. They utilized the inherent anonymity and architectural vulnerabilities of global networks to shift the balance of power without triggering kinetic escalation This continuous, proactive maneuvering was formalized under doctrines of "initiative persistence" and "persistent engagement". These doctrines asserted that digital networks are structurally biased toward the attacker, meaning security cannot be achieved through passive defense. Instead, states are compelled to operate persistently within adversary networks. This allows them to preemptively disrupt malicious capabilities before they can target the domestic homeland.

This historical grey-zone paradigm, however, relied on a critical temporal assumption: that cyber operations occurred at "human speed". Traditional cyber campaigns required human operators to conduct prolonged reconnaissance and manually identify network vulnerabilities. Operators then had to craft highly tailored exploits and execute slow, lateral movements across complex enterprise architectures. . This temporal slack acted as a vital stabilizer in international relations. It provided political leaders, intelligence agencies, and diplomats with a 'pressure release valve.' States could project power, signal displeasure, or gather vital intelligence without forcing adversaries into immediate, escalatory counter-actions. The slow propagation of cyber subversion allowed for human-in-the-loop verification, attribution, and calibrated diplomatic maneuvering, keeping digital competition firmly within the bounds of peacetime statecraft.

The rapid development and deployment of automated, artificial intelligence-driven cyber defense platforms have systematically dismantled this stabilizing temporal slack.Modern enterprise and national-level defensive architectures now operate at 'machine speed.' This is driven by AI-enabled

Extended Detection and Response (XDR) suites, Security Orchestration, Automation, and Response (SOAR) frameworks, and advanced behavioral analysis engines.. These systems ingest and analyze hundreds of millions of data points across global networks in real time, utilizing advanced machine learning algorithms to conduct instant anomaly detection, user entity behavior analysis (UEBA), and automated threat mitigation.

Consequently, the slow, stealthy, and persistent intrusions that once defined grey-zone competition are now instantly detected, mapped, and neutralized by automated defensive agents. AI-defense works by closing the 'Automation Gap,' which is the historical delay between an exploit and the deployment of a patch. By eliminating this buffer, automated systems make low-intensity operations highly visible and immediately ineffective. Rather than allowing cyber operations to remain deniable and slow, AI-defense forces attacks out of the quiet, uncoordinated grey zone and directly into the highly volatile "red zone" of overt, high-stakes kinetic conflict.

This structural collapse of the digital grey zone is severely compounded by the widening "Diplomacy Gap". While national security architectures deploy AI platforms capable of executing automated mitigation decisions in milliseconds, the legal, political, and diplomatic mechanisms of statecraft continue to move at human speed. International law, bilateral treaty negotiations, and formal crisis communication protocols require human deliberation, bureaucratic clearance, consensus-building, and political decision-making—processes that unfold over days, weeks, or months.

An automated AI-defense system might instantly neutralize an adversary's pre-positioned digital access, or misinterpret a routine intelligence probe as an active assault. In either scenario, it executes defensive countermeasures at machine speed without human oversight This creates an extremely dangerous decision-making window. A highly volatile, multi-domain crisis can escalate to the level of armed conflict before a diplomat can draft an email, convene a security council, or establish a direct line of communication with a foreign counterpart.

The primary objective of this research paper is to determine whether the closing of the Automation Gap, far from stabilizing cyberspace, actually destabilizes international relations. This paper evaluates the thesis that defensive AI dominance paradoxically forces rational states into high-intensity, preemptive cycles of aggression. By rendering low-stakes cyber persistence unviable, it accelerates the transition to high-stakes, automated, and highly destructive first-strike cyber-kinetic exploitation.

CHAPTER 2: LITERATURE REVIEW

To comprehend the systemic instability introduced by high-tier AI-cyber defense, international relations (IR) theorists must re-evaluate classic security models under the framework of a "Security Dilemma 2.0". In classical IR theory, the security dilemma is a condition where a state's defensive measures are perceived as offensive threats by its neighbors. For example, physical fortifications or defensive

weaponry are rarely seen as purely protective. Consequently, neighboring states view these developments as highly threatening. Because military hardware is often dual-use, neighboring states feel driven to take preemptive countermeasures. This dynamic initiates a spiral of competitive arms racing. Ultimately, the cycle leaves all participants less secure.

When applied to the digital domain, the deployment of an advanced, automated AI-cyber defense system triggers a highly destabilizing asymmetry. Prominent examples include Singapore's AEGIS-tier national infrastructure and the automated military networks of the United States. A state possessing near-impenetrable, automated digital defenses can operate with strategic impunity. It can execute offensive cyber-kinetic operations or conventional military maneuvers without fear. The state rests secure in the knowledge that its critical infrastructure, military command networks, and financial systems are insulated from digital retaliation This perceived invulnerability destabilizes regional dynamics. Adversaries are forced to abandon low-level digital probing. Instead, they must develop highly aggressive, automated 'first-strike' cyber weapons. These tools are specifically designed to overwhelm, confuse, or physically bypass the defender's AI before it can react. t.

The structural drivers of this dilemma are deeply rooted in the "Automation Gap" and its direct impact on the offense-defense balance. Strategic analysis has long debated whether the architectural properties of cyberspace inherently favor the offensive or defensive actor. Recent theoretical models by Lennart Maschmeyer show that AI integration does not benefit both sides of a cyber conflict equally. Instead, AI automation inherently empowers defense over offense. This asymmetry is defined by a fundamental "Deception-Detection Dichotomy" :

- **Defense is structurally optimized for Detection:** The primary goal of cyber defense is to identify anomalies, recognize malicious patterns, and neutralize unauthorized intrusions as quickly, accurately, and reliably as possible. AI models excel at analyzing massive, high-scale telemetry across complex enterprise networks. They use pattern recognition to identify even the faintest deviations from baseline operations. .
- **Offense is structurally limited by Deception:** The primary goal of cyber offense is to bypass detection, sneak into protected systems, and manipulate those systems to produce a desired, specific effect. This requires creative, novel, and highly adaptive deception. Generative AI models, however, struggle with genuine creativity. They are non-deterministic and prone to hallucinations. Ultimately, they are largely limited to reproducing patterns found within their training datasets.

Cyber operations often transition from basic tasks like mass vulnerability scanning to complex, high-stakes campaigns. These advanced campaigns might include manipulating operational technology or disabling military command networks. As complexity increases, the utility of AI automation for the offense rapidly decreases, while the risk of operational failure grows exponentially.

To demonstrate this dichotomy, Lennart Maschmeyer points to three recent empirical cases that highlight how AI automation operates in real-world cyber conflict :

Comparative Analysis of AI usage in cybersecurity scenarios

Empirical AI Cyber Case	**Core Technical Mechanism**	**Operational Outcome**	**Strategic Implication**
Xbow (The AI "Top Hacker")	AI model topped the leaderboard on HackerOne in mid-2025 by automating vulnerability identification.	Achieved massive *efficiency* in detecting basic, lower-tier ("surface material") vulnerabilities at scale.	Highlighted that AI excels at scale but lacks the creative capability to discover prized, complex zero-days.
2025 Chinese Claude-Automated Attack	State-sponsored group used Anthropic's Claude to automate 80% to 90% of an offensive workflow.	Completely failed against the vast majority of targets; agents relied on known, open-source tools easily detected by defenders.	Confirmed the inherent failure risks of offense automation; AI agent hallucinated credentials and path parameters.
2026 Mexican Government Breach	Small hacktivist group used Claude and ChatGPT to compromise networks and exfiltrate 150 GB of data.	Yielded highly inefficient results compared to traditional, human-led groups (e.g., Chronus, which exfiltrated 2 TB).	Demonstrated that AI automation does not yield a significant increase in offensive effectiveness over manual methods.

These empirical realities confirm that while AI-automated defense can ingest and learn from high-scale data to become increasingly robust, offensive AI agents remain brittle and highly prone to discovery. This dynamic creates a widening "Automation Gap" that favors status-quo powers possessing the immense capital and compute resources required to operate high-scale, data-rich defensive AI. This shift in the offense-defense balance can be mathematically and theoretically modeled through its alteration of the "Subversive Trilemma". This trilemma posits that cyber operations, as instruments of subversion, are fundamentally limited by a negative correlation between speed, intensity, and control. Under this model, an offensive actor cannot maximize all three variables simultaneously. A gain in speed, such as rapid malware propagation, forces a trade-off. It typically results in a loss of control through premature discovery, or a loss of intensity through superficial impacts.

AI-defense systematically exploits these constraints. Automated defenders instantly identify and neutralize known patterns. This forces offensive actors to choose between two flawed options. They can use slow, manual operations to maintain stealth, which sacrifices speed. Alternatively, they can launch volatile, automated attacks, which sacrifices both control and intensity due to the risk of immediate detection.

Paradoxically, this defensive dominance does not lead to strategic peace; instead, it triggers the "Stability-Instability Paradox" in reverse. The paradox was originally conceptualized to explain nuclear deterrence, where stability at the strategic level incentivized low-level proxy conflict beneath the threshold of war. However, in the AI-cyber era, this dynamic inverted.

High-tier AI defenses—such as real-time threat-hunting telemetry, automated vulnerability discovery, and self-healing network protocols—render low-stakes, persistent cyber probing and gradual digital subversion non-viable. Historical paradigms where adversaries maintained multi-year latent access (e.g., classic APT persistence) are effectively eliminated by autonomous defense tools that detect and patch anomalies at machine speed.

Consequently, revisionist actors recognize that grey-zone cyber operations will be neutralized before achieving strategic effect. Rather than disarming, rational states are incentivized to abandon gradual subversion in favor of high-intensity, prompt, and multi-domain operations—such as automated wiper operations, zero-day saturation strikes, or kinetic C4ISR blinding—designed to strike before automated defenses can respond.

Any slow, persistent cyber attempt will be neutralized before it yields strategic value. Consequently, these states are incentivized to bypass the escalation ladder entirely. They shift their behavior from low-stakes cyber persistence to massive, coordinated, and highly destructive "first-strike" cyber-kinetic campaigns. Revisionist states seek to paralyze their targets in a single, decisive blow. They achieve this by executing rapid, machine-speed operations. The goal is to secure a digital fait accompli before the defender's AI can adapt or human diplomats can intervene.

CHAPTER 3: METHODOLOGY

To precisely model how automated offensive agents navigate highly defended networks, computer scientists and strategic planners utilize heuristic search algorithms, most notably Monte Carlo Tree Search (MCTS). When applied to offensive cyber planning, the network environment is modeled as a complex, multi-agent decision space. Within this space, the attacker's objective is to determine the most promising operational path. This involves mapping out lateral movement, credential harvesting, and exploit execution. .An MCTS agent builds a search tree incrementally by executing four distinct phases in an iterative loop: Selection, Expansion, Simulation (or Playout), and Backpropagation. During the Selection phase, the agent navigates down the layers of the tree. It moves from the root node, which represents the initial state of network access, down to an unexpanded leaf node. . This navigation is governed by the Upper Confidence Bound applied to Trees (UCT) formula. The UCT formula mathematically balances a critical trade-off. It weighs 'exploitation,' which means selecting paths known to yield vulnerabilities, against 'exploration,' which involves searching poorly mapped segments of the network.

In a cyberattack scenario, the agent executes the Expansion phase once it selects an unexpanded leaf node. During this phase, it adds one or more child nodes to represent valid next actions. These actions might include scanning a port, executing a local privilege escalation, or harvesting credentials. In the Simulation phase, the agent performs a randomized or heuristic-guided rollout from the newly created node. This rollout continues until it reaches a terminal state, such as a successful domain controller compromise or defensive detection.Finally, during Backpropagation, the resulting payoff of the rollout is propagated back up the ancestral path to the root node. This step updates the $U(n)$ and $N(n)$ values of each node to guide subsequent iterations.

Recent advancements in artificial intelligence have integrated Large Language Models (LLMs) directly into this heuristic framework, creating "LLM-MCTS" architectures. In these systems, the LLM serves dual roles :

- **The World Model:** The LLM provides a "commonsense" prior belief of the network environment, predicting how the target operating systems and defensive configurations will react to specific actions.
- **The Policy Heuristic:** The LLM acts as a selective guide for node expansion, drastically narrowing the search space by proposing only the most logical, semantically coherent, and historically successful exploit chains.

This integration allows automated agents to conduct highly sophisticated, multi-turn 'semantic attacks,' like those modeled in the MUSE-A framework. It also enables on-the-fly decision-making through Reflective MCTS, or R-MCTS. By dynamically reflecting on past failed interactions with network defenses, an R-MCTS agent can adjust its search parameters in real time, bypassing traditional security controls with minimal token usage.

Despite these mathematical efficiencies, offensive MCTS agents suffer from severe structural limitations when confronting real-world, high-stakes targets. The performance of search algorithms degrades exponentially as the width of the search space increases. This issue is highly visible in large enterprise networks containing tens of thousands of endpoints and open ports. This wide search space induces 'tactical blindness.' The MCTS agent fails to identify critical, deep-seated vulnerabilities because its selective node expansion policy overlooks seemingly 'uninteresting' paths. In contrast, a highly creative human hacker would immediately exploit those same pathways.

To illustrate this performance gap, computer scientists compare MCTS against traditional "Sample-and-Filter" (S+F) baselines across different programming languages and data densities.

Furthermore, because LLMs are highly sensitive to minor semantic changes and prone to hallucinations, they frequently attempt to execute non-existent exploits or use hallucinated credentials. This introduces high variance and operational "noise" into the offensive workflow.

When these automated offensive agents (LLM-MCTS) collide with autonomous, national-level defensive AI (AEGIS), they trigger a highly volatile, unmoderated feedback loop. In human-operated campaigns, defenders can observe anomalous activity and execute a calibrated, slow containment strategy. In contrast, the interaction of machine-speed agents occurs in milliseconds. . When an LLM-MCTS agent encounters the rapid, automated firewall rulesets and endpoint isolations of an AEGIS system, it acts immediately. The agent will automatically iterate through its search tree to identify workarounds. The AEGIS system, detecting these rapid, multi-vector probes, escalates its automated response—potentially severing critical dual-use communication links, isolating whole subnetworks, or initiating pre-authorized, automated counter-disruptions. This feedback loop occurs at speeds that completely surpass human cognitive limits. As a result, commanders and policy-makers lose the ability to control, contain, or terminate the escalation. A minor digital incident can easily transform into a severe national crisis in seconds.

CHAPTER 4: CASE STUDY A: The "Persistent Engagement" Model (USA) and Operation Epic Fury / Operation Roaring Lion

The operational limitations of constant presence and the rapid escalation of machine-speed cyber warfare are vividly illustrated by the United States' "Persistent Engagement" doctrine and its deployment in recent conflicts. Persistent Engagement and its diplomatic counterpart, 'Defend Forward,' were formulated by U.S. Cyber Command and enshrined in national cyber strategy. These doctrines mandate that U.S. forces operate continuously in adversary networks. This presence allows them to disrupt malicious activity at its source. However, when adversaries deploy highly integrated, automated AI defenses within their national networks, the doctrine struggles. The constant, intrusive presence required by Persistent Engagement is instantly detected and neutralized. This failure creates severe diplomatic friction and undermines global intelligence collection.

This paradigm shifted dramatically on February 28, 2026, when the United States and Israel launched a massive, coordinated preemptive campaign against Iran, code-named "Operation Epic Fury" by the United States and "Operation Roaring Lion" by Israel. This campaign serves as the premier historical showcase for the integration of kinetic strikes, cyber paralysis, and algorithmic targeting. U.S. Cyber Command acted as the "first mover," initiating cyber disruptions before any physical weapons were deployed. The scale of the cyber offensive was unprecedented: within the opening hours, Israel executed what has been described as the largest cyberattack in global history, collapsing Iran's internet connectivity to a mere 1% to 4% of normal levels. This near-total digital paralysis was achieved through a multi-layered assault. The offensive targeted Border Gateway Protocol (BGP) routing, Domain Name System (DNS) infrastructure, and critical Supervisory Control and Data Acquisition (SCADA) networks. This coordinated action effectively blinded the command-and-control networks of the Islamic Revolutionary Guard Corps (IRGC).

The tactical execution of this joint campaign followed a highly dense, rapid timeline :

Date & Time (2026)	**Operational Phase**	**Key Assets Deployed**	**Strategic & Tactical Targets**
Feb 27	Pre-attack Authorization	Executive order by President Donald Trump.	Sanctioned coordinated military action and pre-emptive cyber disruption.
Feb 28, 1:15 AM EST	Initial Strikes & Cyber Paralysis	100+ US aircraft, B-2 stealth bombers, Tomahawk missiles, USS *Delbert D. Black*.	Coordinated strikes on Pasteur Street district, IRGC Joint HQ, and air defense nodes.
Feb 28, 6:30 AM UTC	Algorithmic Decapitation	Anthropic's Claude LLM, Palantir's Maven Smart System, LUCAS drones.	Fused pattern-of-life and mobile data to target and eliminate Ayatollah Ali Khamenei and IRGC leadership.

Mar 1	Air Superiority & Psych-Ops	Seized state television, hijacked BadeSaba prayer app.	Flooded 5 million devices with defection messages; struck 2,000+ military installations.
Mar 2	Regional Escalation	Kinetic retaliatory strikes by Iranian proxies (Hezbollah, Houthi militias).	Oil infrastructure, maritime transit in Gulf of Oman, and British military base in Cyprus.

The extraordinary tempo of this targeting workflow was enabled by the Pentagon's deployment of advanced artificial intelligence tools. These platforms included Anthropic's Claude large language model and the Palantir-developed Maven Smart System. Embedded across classified intelligence networks, these AI platforms fused massive volumes of signals intelligence (SIGINT), satellite imagery, and human intelligence (HUMINT) at machine speed. The targeting AI tracked Ayatollah Ali Khamenei and over 250 senior IRGC and regime officials with absolute precision. It achieved this by correlating pattern-of-life data, municipal CCTV feeds, mobile networks, and payment systems. This algorithmic targeting facilitated a synchronized, daylight decapitation strike that eliminated much of Iran's top leadership.

The cyber-kinetic targeting chain was further augmented by highly integrated psychological and autonomous operations :

- **Psychological Subversion:** Israeli operators hijacked the BadeSaba prayer calendar application—which possessed over 5 million downloads across Iran—and flooded users' mobile devices with targeted defection messages, while state television networks were seized to broadcast anti-regime messaging.
- **Autonomous Drones:** Low-cost, one-way attack drones from the LUCAS program utilized AI-enabled navigation and terminal guidance to strike localized military infrastructure without human intervention.

The political context of this deployment added a layer of profound strategic irony. The operational integration of Claude occurred just one day after the United States government formally declared Anthropic to be a supply chain risk. Consequently, President Trump had directed government agencies to cease working with the company. This contradiction highlighted how deeply embedded and un-extractable commercial AI architectures have become within sovereign military operational structures.

The strategic fallout of Operation Epic Fury demonstrates how modern hybrid targeting chains are co-dependent and highly escalatory. Deprived of symmetric conventional options, Iran responded with

highly asymmetric, indirect economic warfare.It closed the Strait of Hormuz, which paralyzed 20% of the seaborne oil trade. Additionally, Iran executed kinetic drone and cyberattacks against the energy infrastructure of neutral Gulf states.

Furthermore, the cyber dimension of the conflict created a massive, lingering global exposure. A Microsoft Word N-day vulnerability (CVE-2026-21514) was heavily exploited during the campaign, affecting nearly 14 million assets across targeted countries. Strikingly, the United States carried 99.4% of this exposure, totaling over 15.4 million affected assets across its critical healthcare and government sectors. This footprint leaves these sectors highly vulnerable to pre-positioned access by Iranian state threat groups like MuddyWater and OilRig. These groups are expected to strike once Iran's internet connectivity is restored.

CHAPTER 5: CASE STUDY B: The "Digital Fortress" Model (Singapore) and Operation Cyber Guardian

In stark contrast to the offensive, highly escalatory "Persistent Engagement" model of the United States, Singapore has cultivated a "Digital Fortress" model designed to achieve regional deterrence through sheer defensive resilience, institutional coordination, and transparent capability signaling. Singapore is a highly digitalized, trade-dependent small state balancing major power rivalries in a multipolar Asia. Because of this position, the nation cannot rely on offensive scale or the threat of cyber retaliation. Its posture is anchored in Joseph Nye's framework of deterrence by denial, raising the costs and narrowing the payoff surface for hostile actors attempting to target its critical information infrastructure (CII).

Singapore's defensive architecture is centralized under the Cyber Security Agency (CSA) and the Digital and Intelligence Service (DIS), a dedicated military branch inaugurated in 2022 to protect the sovereign digital domain. A core pillar of Singapore's capability signaling is the Cyber Defence Test and Experimentation Centre (CyTEC). The center was upgraded in 2026. It now leverages advanced AI assistants to simulate highly sophisticated, real-world cyberattack scenarios at scale.

During the annual Critical Infrastructure Defence Exercise (CIDeX 2025), organized jointly by the DIS and CSA, CyTEC's AI tools were utilized to engineer complex attack pathways and intrusion vectors. Over 250 personnel defended Singapore's 11 critical sectors against these live, AI-driven simulated attacks. The exercises specifically targeted 5G telecommunications networks, power grids, and mass transit rail operations. By publicly demonstrating its ability to rapidly contain and recover from systemic digital shocks, Singapore signals "machine-speed resilience" as a credible deterrent.

The operational efficacy of this Digital Fortress model was proven during 'Operation Cyber Guardian.' This national-level defensive mobilization was launched in response to a highly sophisticated intrusion targeting major telecommunications networks. The attack was executed by the state-linked threat actor UNC3886. Known for its ability to operate silently within hypervisors and virtualized environments, UNC3886 routinely bypasses standard security agents, leaving zero-day footprints and evading traditional detection.

Operation Cyber Guardian mobilized over 100 dedicated cyberdefenders from six government agencies to work directly alongside four domestic telecommunications providers. Participating bodies included the CSA, the Infocomm Media Development Authority (IMDA), and the DIS. Rather than relying on standard automated alerts, Singapore's defenders executed a highly coordinated threat-hunting operation. They focused on:

- Centralizing and protecting critical logs.
- Correlating telemetry across multiple disparate systems.
- Monitoring privileged accounts for faint, anomalous behavioral indicators.

Crucially, the operation utilized continuous "purple teaming"—a unified methodology that pairs offensive simulation (Red Teaming) directly with defensive response (Blue Teaming). By executing real-time attacks that mirrored UNC3886's evolving tradecraft, defenders were able to test telemetry, validate controls, and refine threat detections on the fly. Singapore acknowledged the compromise and coordinated a massive, multi-agency response to neutralize the threat before it could transition into systemic sabotage. This successful defense reinforced Singapore's strategic posture: rather than relying on offensive counter-strikes to deter adversaries, the state demonstrates an institutionalized, machine-speed readiness that renders any hostile intrusion strategically futile.

CHAPTER 6: COMPARATIVE ANALYSIS

The stark divergence between the offensive, persistent engagement paradigm of the United States and the defensive, denial-oriented digital fortress paradigm of Singapore highlights the structural fragmentation of modern cyber statecraft. The following comparative matrix delineates these models across their primary strategic, operational, and diplomatic dimensions:

Comparing Cyber Strategic Models:

Strategic Parameter	Persistent Engagement Model (USA)	Digital Fortress Model (Singapore)
Primary Objective	Pre-emptive disruption and seizing of operational initiative outside domestic networks.	Deterrence by denial, systemic resilience, and rapid recovery of critical systems.

Operational Tempo	Machine-speed targeting supported by offensive AI agents and language models.	Continuous, structured monitoring, threat hunting, and automated mitigation.
Institutional Structure	U.S. Cyber Command integrated with national intelligence and conventional military branches.	Centralized coordination via Cyber Security Agency (CSA) and Digital and Intelligence Service (DIS).
Use of Artificial Intelligence	Algorithmic target generation, decision-support wargaming, and autonomous strike navigation.	AI-assisted attack path simulation, digital forensics, and automated anomaly detection.
Diplomatic Signaling	Imposing high costs on adversaries and publicly attributing wrongful acts.	High-profile defense exercises (CIDeX) and regional economic/diplomatic convening.
Geopolitical Posture	Great-power offensive competition and alliance-led "Defend Forward" operations.	Non-aligned, trade-dependent small state balancing major power blocks.

This comparative divergence has birthed a highly destabilizing diplomatic reality: the rise of the "Cyber Fait Accompli". Historically, interstate crises featured a prolonged escalation ladder, allowing diplomats to negotiate, leverage sanctions, or utilize international backchannels to de-escalate tensions. In the era of machine-speed AI targeting and autonomous exploit generation, this critical negotiation phase is bypassed entirely.

A revisionist state can use an LLM-MCTS agent to identify, exploit, and compromise critical infrastructure in milliseconds. By seizing strategic advantage so rapidly, the attacker presents the victim with a completed military reality before its human leadership can even convene. This rapid bypass of diplomacy forces states into a highly reactive posture. Consequently, leaders feel compelled to pre-delegate retaliatory authority to their own autonomous systems. This dynamic significantly increases the probability of accidental, machine-speed escalation.

Furthermore, this technological asymmetry is rapidly eroding the digital sovereignty of small and developing nations, leading to the emergence of rigid "AI-Defense Blocs". Operating high-tier AI-defense platforms requires immense capital, advanced computing hardware, and skilled machine learning talent. Key examples of these systems include AEGIS XDR suites, advanced SIEM architectures, and secure hybrid defense clouds. These resources are highly concentrated within a small number of tech-superpowers, primarily the United States and China.

Small and medium-sized states rarely have the capital to develop sovereign, AEGIS-tier systems. As a result, they face a coercive choice. They can remain vulnerable to low-cost, automated cybercrime and state disruption, or they can surrender their digital sovereignty by joining an AI-defense bloc led by a superpower. To get protection, these states must adopt foreign, black-box AI security architectures and route their network telemetry through superpower data centers. Consequently, these vassal states lose the ability to independently govern or secure their own digital infrastructure. This dependency further polarizes the global geopolitical landscape

In response to these threats, the United Nations has sought to institutionalize and formalize global cyber governance. On March 30–31, 2026, representatives from all 193 UN Member States convened in New York to launch the first permanent Global Mechanism on Information and Communication Technology (ICT) Security. This historic forum serves as the permanent successor to decades of ad hoc, temporary frameworks, most notably the Groups of Governmental Experts (UN GGE, 2004–2021) and the Open-ended Working Groups (UN OEWG, 2019–2025).

Ambassador Egriselda López of El Salvador chairs the new Global Mechanism. The body is designed to promote responsible state behavior and implement the 11 voluntary cyber norms agreed upon by member states. Additionally, it establishes Dedicated Thematic Groups (DTGs) to address substantive threats and capacity building.

Despite the diplomatic consensus underpinning its launch, the Global Mechanism immediately exposed deep, structural fault lines that reflect a fundamental geopolitical divide. On one side, Russia and China spearhead a state-centric vision of 'information sovereignty.' This model seeks to restrict cross-border data flows, tighten administrative control over domestic networks, and criminalize online political dissent. On the other side is a Western, rules-based vision of an open, free, and secure global cyberspace. This ideological divide renders even basic procedural matters—such as balancing consensus-driven decision-making with operational efficiency—deeply political, slowing the implementation of cyber norms.

Furthermore, multilateral governance efforts are structurally constrained by a "Legal Trilemma," consisting of three distinct failure modes that render traditional international law ineffective against modern digital harms.

To analyze these failure modes, international lawyers contrast their specific technical mechanisms and real-world geopolitical impacts :

Legal Failure Mode	Core Technical Mechanism	Real-World Geopolitical Impact	Evasive Countermeasure
The Classification Gap	Legal categories fail to keep pace with rapid technological evolution.	Courts spend years debating basic definitions (e.g., whether cryptocurrency is "property").	Decentralized finance (DeFi) architectures engineered to bypass judicial asset freezing.
The Implementation Gap	Enforcement protocols move at human speed while asset flows move at machine speed.	Mutual legal assistance treaties (such as the Hanoi Convention) require weeks to trace assets.	Assets swapped into privacy coins (Monero) and obscured within minutes of discovery.
The Attribution Gap	Harms originate from sovereign territories that refuse to cooperate with investigators.	Host states are not merely uncooperative but are, in some cases, structurally implicated in the crime.	Placement of physical infrastructure and personnel within non-cooperative jurisdictions.

This legal trilemma is vividly illustrated by the rise of industrial-scale cryptocurrency investment fraud, commonly known as "pig butchering". Operating within specialized, highly fortified compounds concentrated in non-cooperative jurisdictions like Myanmar, Cambodia, and Laos, these networks extracted an estimated $9.9 billion in 2024 alone. The compounds are staffed by hundreds of thousands of human trafficking victims forced to execute sophisticated social engineering campaigns under threat of severe physical violence.

Because the host states are non-parties or have failed to ratify international treaties like the Budapest and Hanoi Conventions, they structurally refuse to cooperate with international investigators. The pig butchering compound architecture was specifically engineered to exploit gaps in international law. Operators route financial flows through unregulated cryptocurrency rails and communicate via encrypted platforms. Furthermore, they base their physical operations in sovereign, uncooperative territories. Consequently, while the legal classification of the fraud is clear, the international community possesses zero viable pathways to enforce the law, demonstrating the profound limits of multilateral cyber norms.

CHAPTER 7: DISCUSSION

The structural instability introduced by high-tier AI-cyber defenses reaches its most dangerous inflection point when interacting with dual-use military systems. In modern military architectures, tactical conventional systems overlap dangerously with strategic nuclear networks. Specifically, conventional forces share critical Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) infrastructure. The same communication arrays, satellite links, and sensor networks that enable conventional precision-guided weapons are also utilized to transmit nuclear strategic warning and launch authorization. This convergence completely erodes the traditional structural "firebreaks" that separated conventional and strategic nuclear escalation.

When a state deploys an automated, AI-driven targeting and cyber exploitation platform (such as the Maven or Claude-supported targeting chains used in Operation Epic Fury) during a conventional regional conflict, the AI is optimized to compress the kill chain by rapidly identifying and disabling the adversary's C4ISR nodes. Conventional sensors and command networks directly overlap with strategic nuclear systems. Consequently, an automated cyber assault designed to blind air defenses is indistinguishable from the opening stage of a nuclear first strike.

This ambiguity triggers extreme, automated "use-it-or-lose-it" pressures—forcing a state to either use its strategic nuclear assets immediately or risk losing its command authority permanently. Under this pressure, a nuclear-armed state will perceive the machine-speed blinding of its command networks as an existential threat. This is especially true for states like China, North Korea, or Pakistan, which may lack confidence in the survivability of their second-strike capabilities. Digital paralysis prevents national leaders from verifying the scope of an attack at human speed. As a result, leadership faces an agonizing choice: launch their strategic nuclear forces before communications are severed ("use it"), or risk having their nuclear deterrent neutralized by automated systems ("lose it").

Thus, the pursuit of machine-speed "decision superiority" through automated cyber-targeting directly increases the risk of catastrophic, inadvertent nuclear escalation.

To mitigate these profound structural risks and restore strategic stability to the international system, policy-makers, computer scientists, and diplomatic delegations must coordinate to implement three targeted, non-negotiable interventions:

1. Cryptographic Machine-Speed Hotlines

Traditional voice-based or human-to-human crisis hotlines are entirely inadequate to manage conflicts unfolding in milliseconds. The international community must develop automated, cryptographic "machine-speed hotlines" integrated directly into sovereign national AI-defense systems. These secure, pre-negotiated protocols would allow national AI platforms to exchange structured, cryptographic tokens of intent during high-velocity network events. These systems automatically signal that a network probe or remediation is localized and non-strategic. In doing so, they bypass human latency and defuse unmoderated feedback loops. This automation prevents systemic misunderstandings before they trigger kinetic retaliation.

2. Mandatory Human-in-the-Loop Agentic AI Guardrails

Governments and defense contractors must refuse to deploy fully autonomous agentic AI in offensive or high-risk defensive cyber-targeting applications. Planners should formally adopt information security frameworks, such as the Forrester Agentic AI Guardrails (AEGIS) Framework, which map directly to the joint security guidance issued by the Five Eyes cybersecurity agencies. These frameworks must encode "Human-in-the-Loop" requirements not as mere policy recommendations, but as hardcoded, cryptographically enforced approval gates. Explicit, authenticated human authorization must be required for any action with irreversible downstream effects. Examples include disabling dual-use communications, manipulating SCADA(Supervisory Control and Data Acquisition) systems, or executing BGP(Border Gateway Protocol) hijacks. Enforcing this step preserves the temporal space necessary for crisis diplomacy.

3. Decentralized AI-Defense Capacity-Building

To prevent the coercive formation of destabilizing "AI-Defense Blocs" and preserve the digital sovereignty of small and developing nations, the UN Global Mechanism must prioritize multilateral technology transfer and capacity-building. Utilizing the Global Mechanism's Dedicated Thematic Group on capacity building, developed nations must share standardized, open-source, and verifiable AI-defensive models with developing states. The international community must democratize access to robust, automated detection capabilities. Crucially, this can be done without forcing small states to surrender their network telemetry to foreign superpowers. This approach flattens the technological hierarchy, reduces global vulnerability, and stabilizes the digital domain.

REFERENCES

Triglianos, Charles and Evans, Luke. "Lumiere Outline: How the Automation Gap and AI-Defense Destabilize Strategic Relationships." *Lumiere Technical Outline PDF*, May 2026.

The Forge (Defence Australia). "How Can Small States Deter Cyberattacks? Estonia and Singapore in Comparative Perspective." *The Forge*, March 2025. URL: https://theforge.defence.gov.au/article/how-can-small-states-deter-cyberattacks.

Harknett, Richard J. "The Escalation Inversion and Other Oddities of Situational Cyber Stability." *Texas National Security Review*, Vol. 3, No. 4, Autumn 2020. URL: https://tnsr.org/2020/09/the-escalation-inversion-and-other-oddities-of-situational-cyber-stability/.

Lindsay, Jon R., and Gartzke, Erik. "Coercion Through Cyberspace: Strategic Logic and Empirical Realities." UC San Diego Policy Brief, Draft Series, 2022. URL: https://deterrence.ucsd.edu/_files/LindsayGartzke_CoercionThroughCyberspace_DraftPublic1.pdf.

Harknett, Richard J. "Setting the Stage: Cyber Contingency Campaigning and the Logic of Initiative Persistence." *Lawfare*, July 14, 2022. URL: https://www.lawfaremedia.org/article/setting-the-stage--cyber-contingency-campaigning.

Smeets, Max. "US Cyber Strategy of Persistent Engagement and Defend Forward." Center for Security Studies (CSS) ETH Zürich, February 2020. URL: https://css.ethz.ch/en/center/CSS-news/2020/02/us-cyber-strategy-of-persistent-engagement-defend-forward.html.

Nakasone, Paul M. "Leaders Discuss DoD's Cyber Strategy to Protect America and Partners." *War.gov*, May 2018. URL: https://www.war.gov/News/News-Stories/Article/Article/2862784/leaders-discuss-dods-cyber-strategy-to-protect-america-partners/.

Royal United Services Institute (RUSI). "Brief, Bold, and Beautiful: Reactions to the US National Cyber Strategy." *RUSI Commentary*, March 2026. URL: https://www.rusi.org/explore-our-research/publications/commentary/brief-bold-and-beautiful-reactions-us-national-cyber-strategy.

Maschmeyer, Lennart. "Subversion, Cyber Operations, and Reverse Structural Power in World Politics." *European Journal of International Relations*, Vol. 28, No. 3, August 2022. URL: https://www.research-collection.ethz.ch/server/api/core/bitstreams/9c0c24a1-43a1-4695-87eb-6ffc1c7ba91e/content.

Healey, Jason. "Cyber Operations and Maschmeyer's 'Subversive Trilemma'." *Lawfare*, July 14, 2022. URL: https://www.lawfaremedia.org/article/cyber-operations-and-maschmeyers-subversion-trilemma.

Maschmeyer, Lennart. "The AI Revolution in Cyber Conflict: Why Offense Automation Falls Short." *Lawfare*, April 8, 2026. URL: https://www.lawfaremedia.org/article/the-ai-revolution-in-cyber-conflict.

AI Aegis Lab. "AEGIS Extended Detection and Response (XDR) Platform Security Suite." *AI Aegis Lab Product Guide*, 2025. URL: https://aiaegislab.com/.

Aegis Cyber Defense Systems. "Aegis Defender Pro and Master Block List Autonomous Security Analytics." *Aegis CDS Product Reference*, 2026. URL: https://aegiscds.com/company/about.

The Hague Program on International Cyber Security. "Launching the UN Global Mechanism on Cybersecurity: Three Ways Technology Breaks the Law." *Cyber Policy Review*, April 2026. URL: https://policyreview.info/articles/news/launching-un-global-mechanism.

Meyer, Paul. "Global Cyber Security – At Last a 'Permanent Mechanism' at the UN." *ICT4Peace*, January 2026. URL: https://ict4peace.org/activities/global-cyber-security-at-last-a-permanent-mechanism-at-the-un/.

The Forge (Defence Australia). "Autonomous and Automated Weapon Systems and Traditional Dynamics of Conflict Escalation." *The Forge*, February 2025. URL: https://theforge.defence.gov.au/article/autonomous-and-automated-weapon-systems-and-traditional-dynamics-conflict-escalation.

Johnson, James. "The End of Strategic Stability? Military AI, Command and Control, and Crisis Instability." *The Washington Quarterly*, Vol. 43, No. 3, September 2020. URL: https://doras.dcu.ie/25512/1/TWQ%20JamesJohnson%20(2020).pdf.

The Hague Centre for Strategic Studies (HCSS). "Arms Control: Shifting Sands of Strategic Stability." *HCSS Research Report*, February 2022. URL: https://hcss.nl/wp-content/uploads/2022/02/Arms-Control-Shifting-sands-of-strategic-stability-HCSS-2022.pdf.

Hersman, Rebecca, and Younis, Reja. "Surveillance, Situational Awareness, and Warning at the Conventional-Strategic Interface." *Nuclear Network CSIS*, January 15, 2021. URL: https://nuclearnetwork.csis.org/surveillance-situational-awareness-and-warning-at-the-conventional-strategic-interface/.

Center for Strategic and International Studies (CSIS). "An Adversary Gets a Vote: Information Dominance and Strategic Stability risks." *CSIS Analysis*, July 2021. URL: https://www.csis.org/analysis/adversary-gets-vote.

Maschmeyer, Lennart. "The AI Revolution in Cyber Conflict." *Lawfare*, April 8, 2026. URL: https://www.lawfaremedia.org/article/the-ai-revolution-in-cyber-conflict.

Maschmeyer, Lennart. "Deception and Detection: Why Artificial Intelligence Empowers Cyber Defense over Offense." *International Security*, Vol. 50, No. 3, Winter 2025/26, pp. 86–126. doi: 10.1162/ISEC.a.398.

Maschmeyer, Lennart. "The Subversive Trilemma: Why Cyber Operations Fall Short of Expectations." *International Security*, Vol. 46, No. 2, Fall 2021, pp. 51–90. doi: 10.1162/isec_a_00418.

Maschmeyer, Lennart. "The Myth of Cyberwar and the Realities of Subversion." *Modern War Institute at West Point*, November 2021. URL: https://mwi.westpoint.edu/the-myth-of-cyberwar-and-the-realities-of-subversion/.

Maschmeyer, Lennart. "The Subversive Trilemma of Cyber Operations: Covert Action and Power Projection." *Waterloo Cybersecurity Seminar Series*, February 2022. URL: https://uwaterloo.ca/cybersecurity-privacy-institute/cpi-talk-subversion-spies-hackers.

Maschmeyer, Lennart. "Subversion, Cyber Operations, and Reverse Structural Power." *European Journal of International Relations*, August 2022. URL: https://www.research-collection.ethz.ch/bitstreams/1267d20d-6a50-4ebd-b478-640c02d2792b/download.

The Friday Times. "Modern War: Perception Strikes First." *The Friday Times*, May 11, 2026. URL: https://www.thefridaytimes.com/11-May-2026/modern-war-perception-strikes-first.

Hersman, Rebecca. "Surveillance, Situational Awareness, and Warning at the Conventional-Strategic Interface." *Nuclear Network CSIS*, January 15, 2021. URL: https://nuclearnetwork.csis.org/surveillance-situational-awareness-and-warning-at-the-conventional-strategic-interface/.

Wikipedia. "Monte Carlo Tree Search." *Wikipedia, The Free Encyclopedia*, 2026. URL: https://en.wikipedia.org/wiki/Monte_Carlo_tree_search.

Roelofs, G. "Pitfalls and Solutions When Using Monte Carlo Tree Search for Strategy and Tactical Games." *Game AI Pro 3*, Chapter 28, CRC Press, 2017. URL: http://www.gameaipro.com/GameAIPro3/GameAIPro3_Chapter28_Pitfalls_and_Solutions_When_Using_Monte_Carlo_Tree_Search_for_Strategy_and_Tactical_Games.pdf.

UC Berkeley CS188. "Games: Monte Carlo Tree Search." *CS188 Intro to AI Textbook*, Fall 2024. URL: https://inst.eecs.berkeley.edu/~cs188/textbook/games/monte-carlo.html.

IEEE Xplore. "Design Methodology for an Automated Penetration Testing System Based on Improved Monte Carlo Tree Search." *IEEE Conference Publication*, October 2023 / April 2025. doi: 10.10965407.

IEEE Xplore. "Design Methodology for an Automated Penetration Testing System Based on Improved Monte Carlo Tree Search." *IEEE Conference Publication*, October 2023 / April 2025. doi: 10.10965407.

Sutton, Richard, and Barto, Andrew. "Monte Carlo Tree Search - Single-Agent MDPs Overview." *Reinforcement Learning Notes*, University of Queensland Press, 2023. URL: https://gibberblot.github.io/rl-notes/single-agent/mcts.html.

Hutter, Marcus. "The Animated Monte Carlo Tree Search (MCTS)." *Medium Data Science*, February 2024. URL: https://medium.com/data-science/the-animated-monte-carlo-tree-search-mcts-c05bb48b018c.

NeurIPS. "Planning with Large Language Models for Code Generation: LLM-MCTS." *Conference on Neural Information Processing Systems (NeurIPS) Proceedings*, Vol. 36, December 2023. URL: https://proceedings.neurips.cc/paper_files/paper/2023/hash/65a39213d7d0e1eb5d192aa77e77eeb7-Abstract-Conference.html.

ACL Anthology. "MUSE: Multi-turn Safety Alignment Framework (MUSE-A Framework)." *Empirical Methods in Natural Language Processing (EMNLP) Main Volume 2025*, pages 21282--21303. URL: https://aclanthology.org/2025.emnlp-main.1080/.

Arxiv. "MUlti-turn SEmantic Attack (MUSE-A): Frame Semantics and MCTS." *ArXiv Preprint*, Vol. 2509.14651v1, September 2025. URL: https://arxiv.org/html/2509.14651v1.

OpenReview. "Reflective Monte Carlo Tree Search (R-MCTS) for AI Agent Decision Space." *OpenReview Board Forum*, ID: GBIUbwW9D8, April 2026. URL: https://openreview.net/forum?id=GBIUbwW9D8.

Roelofs, G. "Pitfalls and Solutions When Using Monte Carlo Tree Search for Strategy and Tactical Games." *Game AI Pro 3*, Chapter 28, CRC Press, 2017. URL: http://www.gameaipro.com/GameAIPro3/GameAIPro3_Chapter28_Pitfalls_and_Solutions_When_Using_Monte_Carlo_Tree_Search_for_Strategy_and_Tactical_Games.pdf.

Patro, Arun. "LLM Generation Optimization: Code Generation Pass Rate of Sample-and-Filter vs. Monte Carlo Tree Search." *Arun Patro Blog*, November 2024. URL: https://arunpatro.github.io/blog/mcts/.

Worthington, Janet; Cairns, Geoff; Pollard, Jeff; Iannopollo, Enza; and Budge, Jinan. "Careful Adoption of Agentic AI Services Guidance and Forrester's AEGIS Framework." *Forrester Research Blogs*, May 1, 2026. URL: https://www.forrester.com/blogs/five-eyes-cybersecurity-agencies-careful-agentic-ai-adoption-guidance-operationalized-by-aegis/.

Smeets, Max. "US Cyber Strategy of Persistent Engagement and Defend Forward." Center for Security Studies (CSS) ETH Zürich, February 2020. URL: https://css.ethz.ch/en/center/CSS-news/2020/02/us-cyber-strategy-of-persistent-engagement-defend-forward.html.

Radware Threat Intelligence. "Timeline of Preemptive Strikes and Regional Responses: Operation Epic Fury." *Radware Security Advisory*, March 2026. URL: https://www.radware.com/security/threat-advisories-and-attack-reports/ddos-activity-following-operation-epic-fury-roaring-lion/.

AttackIQ. "Situation Report: Operation Epic Fury and Allied Response." *AttackIQ Assessment Reports*, March 5, 2026. URL: https://www.attackiq.com/2026/03/05/operation-epic-fury/.

Radware Threat Intelligence. "Timeline of Preemptive Strikes and Regional Responses: Operation Epic Fury." *Radware Security Advisory*, March 2026. URL: https://www.radware.com/security/threat-advisories-and-attack-reports/ddos-activity-following-operation-epic-fury-roaring-lion/.

Homeland Security Today. "Algorithmic Warfare in the Iran Conflict: Operation Epic Fury / Operation Roaring Lion." *HSA Today*, March 6, 2026. URL: https://www.hstoday.us/subject-matter-areas/ai-and-advanced-tech/algorithmic-warfare-in-the-iran-conflict-operation-epic-fury-and-dawn-of-the-ai-battlefield/.

Irregular Warfare Center. "Operation Epic Fury / Roaring Lion Is an Irregular Warfare Showcase." *IWC Insights*, mid-April 2026. URL: https://irregularwarfarecenter.org/publications/insights/operation-epic-fury-roaring-lion-is-an-iw-showcase/.

Washington Examiner. "Israel's Cyber Decapitation Operation Exposed the Mullahs' Paper Tigers." *Washington Examiner Op-Ed*, February 2026. URL: https://www.washingtonexaminer.com/op-eds/4566219/israel-cyber-decapitation-exposed-iran-mullahs-paper-tigers/.

Cyble. "The Tripartite Cyber War: Coordinated Cyber Warfare at Global Scale Analysis." *Cyble Threat Reports*, April 2026. URL: https://cyble.com/resources/research-reports/iran-israel-us-cyber-warfare/.

Center for Strategic and International Studies (CSIS). "How Will Cyber Warfare Shape the US-Israel Conflict with Iran?" *CSIS Q&A Analysis*, March 2026. URL: https://www.csis.org/analysis/how-will-cyber-warfare-shape-us-israel-conflict-iran.

Chatham House. "The Iran War Highlights the Creeping Use of AI in Target Generation." *Chatham House Journal*, March 27, 2026. URL: https://www.chathamhouse.org/2026/03/iran-war-highlights-creeping-use-ai-warfare.

Tenable. "Operation Epic Fury: Why Exposure Data Rebalances the Threat Picture." *Tenable Technical Blogs*, March 2026. URL: https://www.tenable.com/blog/operation-epic-fury-why-exposure-data-changes-everything-about-irans-cyber-kinetic-campaign.

Singapore Armed Forces (SAF). "The Digital and Intelligence Service (DIS) Branch Inauguration and SAF Structure." *Wikipedia, The Free Encyclopedia*, 2026. URL: https://en.wikipedia.org/wiki/Singapore_Armed_Forces.

Ministry of Defence (MINDEF) Singapore. "Fact Sheet: Upgraded Cyber Defence Test and Experimentation Centre (CyTEC) and the SAF Digital Range." *MINDEF Singapore Releases*, February 27, 2026. URL: https://www.mindef.gov.sg/news-and-events/latest-releases/27feb26-fs2/.

Cyber Security Agency of Singapore (CSA). "AI-Enabled Capabilities and Industry Collaboration Elevate Defense at National Cyber Defence Exercise (CIDeX 2025)." *CSA Press Releases*, November 12, 2025. URL: https://www.csa.gov.sg/news-events/press-releases/ai-enabled-capabilities-and-collaboration-with-industry-elevate-defence-of-critical-infrastructure-at-national-cyber-defence-exercise/.

Ministry of Defence (MINDEF) Singapore. "Critical Infrastructure Defence Exercise co-organised by DIS and CSA supported by iTrust SUTD." *MINDEF News Release*, November 12, 2025. URL: https://www.mindef.gov.sg/news-and-events/latest-releases/12nov25-nr2/.

Singapore University of Technology and Design (SUTD). "CIDeX 2025 Incorporates Real Combat Intelligence for Realistic Training." *SUTD News Listing*, November 14, 2025. URL: https://www.sutd.edu.sg/news-listing/cyber-defence-exercise-incorporates-real-combat-intelligence-for-realistic-training/.

Singapore Cyber Security Resources. "UNC3886 Telco Defence Effort: Operation Cyber Guardian." *IMDA & DIS Coordinated Actions Report*, April 2026. URL: https://www.privacy.com.sg/resources/unc3886-telco-defence-effort/.

International Institute for Strategic Studies (IISS). "Cloud Adoption for National Security and Defence Purposes: Hybrid Cloud Integration in Singapore." *IISS Research Paper Series*, Pub 25-111, December 2025. URL: https://www.iiss.org/globalassets/media-library---content--migration/files/research-papers/2025/12/cloud-adoption-for-national-security-and-defence-purposes/pub25-111---aws-transitioning-to-cloud-v6-2.pdf.

Hoover Institution. "The AI Titans' Security Dilemmas: Strategic Competition and Military Intelligentization (智能化)." *Hoover Institution Press*, April 2026. URL: https://www.hoover.org/research/ai-titans.

Taylor & Francis. "Voluntary Cyber Norms and State Behavior in Cyberspace: From UN GGE to UN OEWG." *Journal of Cybersecurity and International Law*, Vol. 3, No. 1, March 2026. doi: 10.1080/23738871.2026.2649932.

Meyer, Paul. "Global Cyber Security – At Last a 'Permanent Mechanism' at the UN." *ICT4Peace*, January 2026. URL: https://ict4peace.org/activities/global-cyber-security-at-last-a-permanent-mechanism-at-the-un/.

UNIDIR. "Geneva Cyber Week 2026: Cyber Stability Conference (CS26) Programme." *United Nations Institute for Disarmament Research*, April 2026. URL: https://unidir.org/wp-content/uploads/2026/05/UNIDIR_CS26_Programme-1.pdf.

UNIDIR. "Engaging Regions: Insights into the Global Mechanism on ICT Security." *UNIDIR Event Series*, March 2026. URL: https://unidir.org/event/engaging-regions-insights-into-the-global-mechanism-on-ict-security/.

Ministry of Foreign Affairs of Japan. "United Nations Global Mechanism Establishment Meeting on Cybersecurity." *MOFA Japan Press Releases*, April 1, 2026. URL: https://www.mofa.go.jp/press/release/pressite_000001_02230.html.

Permanent Mission of El Salvador to the UN. "The New United Nations Permanent Mechanism on Cybersecurity and Dedicated Thematic Groups (DTGs)." *UN Interface Publications*, March 2026. URL: https://www.interface-eu.org/publications/the-new-united-nations-mechanism-on-cybersecurity.

UN Interface Publications. "From Launch to Delivery: The UN's New Cybersecurity Mechanism Faces Old Divides." *UN Security Negotiations Briefing*, March 2026. URL: https://www.interface-eu.org/publications/from-launch-to-delivery-the-uns-new-cybersecurity-mechanism-faces-old-divides.

Maschmeyer, Lennart, and Lindsay, Jon. "Nuclear Command and Control Vulnerability under the Cyber Shadow." *Journal of Cybersecurity*, Oxford Academic, Vol. 3, No. 1, November 2020. URL: https://academic.oup.com/cybersecurity/article/3/1/37/2996537.

Hersman, Rebecca, and Younis, Reja. "Surveillance, Situational Awareness, and Warning at the Conventional-Strategic Interface." *Nuclear Network CSIS*, January 15, 2021. URL: https://nuclearnetwork.csis.org/surveillance-situational-awareness-and-warning-at-the-conventional-strategic-interface/