Artificial Intelligence and Independent Cybersecurity

(Translated)

https://www.al-waie.org/archives/article/20482

Al Waie Magazine Issue No. 482

Forty first Year, Rabi' I 1448 AH corresponding to August 2026 CE

Abdel-Halim Al-Hourani

I write these lines knowing that much of what follows will soon seem familiar; the pace of change in this field has outstripped our ability to document it. The issue is no longer merely a subject of research; artificial intelligence has actively entered the realm of cyber warfare and permeated the details of our daily lives. A state that lags in building its defenses today stands to lose not just a system or two, but its entire margin of maneuverability—precisely at the moment it needs to exercise its sovereign decision-making power most. While preparing entails high costs, the regret of failing to prepare will prove far costlier when the time comes.

How Has Artificial Intelligence Changed the Rules of the Game?

There is a common confusion regarding the different types of artificial intelligence that affects our understanding of the threats involved. To simplify this for the general public, we can distinguish between them as follows:

- **Discriminative AI:** This is the oldest and most prevalent form; its function is to sort and classify items. It operates much like email filters that automatically identify and categorize spam messages.

- **Generative AI:** This is the type seen in modern models like ChatGPT; it is capable of creativity and generating entirely new content based on the data upon which it was trained.

Before the advent of generative models, cyberattacks required organized teams and substantial budgets for reconnaissance, keeping offensive capabilities largely in the hands of major powers and a select few groups. Today, however, artificial intelligence has drastically reduced the cost of offensive operations, enabling small groups or individuals with moderate skills to inflict serious damage at a negligible cost. Key Offensive Shifts:

- **Smart Reconnaissance:** Tasks that once required weeks of human effort to gather intelligence and map out a target are now performed by digital agents in mere hours—surpassing current defensive capabilities.

- **Automated Vulnerability Generation:** AI tools scan software for vulnerabilities and convert them into actual exploit code in record time, outpacing the target's ability to patch their systems.

- **Smart Spear-Phishing:** The era of clumsy phishing emails is over; AI analyzes a victim's digital footprint and communication style to craft personalized messages that are difficult to detect, even for seasoned professionals.

- **Deepfakes:** Moving beyond mere demonstrations, a short audio sample is now sufficient to mimic a voice with high fidelity, enabling the creation of fake audio or video conversations to deceive victims and persuade them to authorize massive financial transfers.

- **Polymorphic Malware:** Malicious software that alters its form and code with every transmission, confounding traditional detection systems and allowing the virus to persist within the system undetected.

Attacks on the "Brain" of AI:

A fundamental shift has occurred: the target is no longer just network systems, but the AI "brain" itself. These attacks take three primary forms:

1. Deliberately Feeding AI Incorrect Information (Data Poisoning): Attackers inject misleading information during the training phase, causing the system to adopt flawed logic

without detection. An example is a construction contract auditing system secretly trained to ignore delay penalties in specific contracts.

2. Misleading the System With Clever, Fabricated Inputs (Deception Attacks): After training is complete, an attacker introduces precisely modified inputs to fool the model—such as a subtle tweak to a virus file that causes the security system to read it as a standard text file, or an imperceptible alteration to an image that confuses a facial recognition system.

3. Stealing System Secrets and Internal Data (Model Extraction): Retrieving sensitive data or stealing the algorithm itself; if a system has been trained on a company's trade secrets, an attacker might use clever queries to trick it into revealing that confidential information.

Critical Sectors in the Eye of the Storm: What Does the Disaster Look Like?

Not all systems are of equal value; critical national infrastructure is the top priority for protection. Here are four realistic scenarios illustrating how artificial intelligence can transform the nature of the threat:

Scenario One: Collapse of the Electrical Grid and Energy Sector

An AI-coordinated attack targets multiple substations simultaneously by exploiting vulnerabilities in industrial control systems. Within minutes, cascading failures begin; within hours, the country is left without power, and backup batteries in hospitals and data centers are depleted. The disaster lies not in the darkness itself, but in the immediate, total paralysis that follows as communications, banking, and electronic payment systems fail. In this AI-enhanced scenario, the conflict is not "point-versus-point"—as seen in Ukraine in 2015—but rather "algorithm-versus-algorithm," where the attack coordinates multiple strikes and adapts to defensive measures in real time.

Scenario Two: Paralysis of the Banking and Financial System

A multi-layered attack begins with a deepfake call from a high-ranking financial official, coinciding with ransomware striking the banking settlement system. Within hours, instant payments cease; within a day, people queue at ATMs for mass cash withdrawals, while automated accounts spread rumors that undermine confidence in the national currency. This mirrors the 2022 incident in Costa Rica, where a ransomware attack paralyzed state institutions and forced a declaration of emergency; however, with the advent of AI, the situation escalates from operational paralysis to a full-blown crisis of monetary sovereignty.

Scenario Three: Erosion of Social Trust and the Information Space

This is the most dangerous scenario, as it requires no technical breach but instead exploits the open nature of the internet. AI generates thousands of coordinated automated accounts that disseminate high-fidelity deepfake videos of officials and articles mimicking established media outlets. The goal is not to convince everyone, but to sow widespread doubt, leaving citizens unable to distinguish between the real and the fake. This paralyzes the state's ability to manage crises, as every official statement is met with skepticism. Researchers call this the "liar's dividend," and it carries the highest probability of occurrence.

Scenario Four: Hijacking Weaponry and Defense Systems

With the proliferation of drones and autonomous systems, a threat emerges regarding the targeting of control systems to disable a drone fleet or turning it against its owner. This requires compromising the software supply chain or poisoning target-identification data. Several fronts have witnessed incidents of navigation system jamming that caused drones to turn back against friendly forces; furthermore, as AI-driven attacks evolve, it becomes possible for a friendly radar to misidentify a friendly asset as an enemy, or vice versa.

Can a State Truly Collapse?

The answer depends on the type of collapse:

· **Sudden, Total Collapse:** Unlikely; modern states possess multiple redundancies that prevent sectors from failing all at once. Military forces operate on isolated networks, and critical infrastructure systems often have manual override capabilities.

· **Gradual, Multi-Stage Paralysis:** Far more likely; a coordinated attack striking three or four sectors in rapid succession—coupled with a disinformation campaign—can push a state to the point of losing effective control. The difference between the two is akin to the difference between an explosion and a slow, corrosive fire: a state might recover from a physical earthquake in months, whereas a "cyber-earthquake" that shatters trust takes years to heal.

The Triple Firewall: How Can a State Protect Its Independence and Decision-Making Autonomy?

Effective defense is not built from disjointed measures but from an integrated system comprising three fundamental pillars:

1. Governance and Legislation (Regulation)

Establishing a cybersecurity and artificial intelligence authority with independent management, a subordinate cyber command, and a direct reporting line to the state's highest leadership. Its mandate would be to enact binding regulations and directives for sensitive sectors, establishing minimum standards and control measures.

2. Independence Technical Capabilities (The Machine)

Independence entails full control over critical nodes by establishing an independent cloud infrastructure for storing and processing sensitive government data, and by building specialized computing centers and foundational AI models trained locally on independent data.

3. Human Capital (The Human)

Technology without a workforce is an empty shell. A successful plan places humans at the forefront by preparing highly skilled, specialized teams and research centers capable of producing an army of professionals across diverse disciplines—including intelligence, engineering, psychology, economics, and political science—to create an integrated defense and offense ecosystem.

The Army of Defensive Systems and the State Protocol

The concept of an "AI Defense Army" stems from the reality that a high-speed machine cannot be countered by slow human minds alone; attacks occur at lightning speed, and personnel behind screens cannot repel them. Instead, we require defensive "automation" capable of responding in fractions of a second. This framework comprises three components:

First: Intelligent Digital Systems

· **Intelligent Operations Center:** A monitoring hub that filters out thousands of trivial alerts and automatically resolves minor issues, while referring major, sensitive decisions to human managers.

· **Proactive Hunter:** Deploying digital agents to scour the system for any hidden traces of intruders before an attack begins; prevention is less costly than remediation.

· **Ethical Offensive Testing Teams:** Software and hackers tasked with continuously attacking state systems to identify and patch vulnerabilities before adversaries can exploit them.

Second: Digital Independence and Content Protection

· **Digital Sealing:** Applying an invisible digital authentication mark to official data to protect society from rumors and information ambiguity; any statement lacking this seal is deemed fake.

· **Technical Independence:** Developing indigenous artificial intelligence that understands our culture and safeguards state data privacy, rather than relying on external companies that might leak data.

Third: State and Individual Strategy

· **Comprehensive Cyber Exercises:** Testing the response of the entire government—not just technical teams—to identify gaps and train leaders in managing digital emergencies.

· **Critical Business Continuity Plans:** An annually tested plan for every sector, covering scenarios such as total system failure, the shift to manual operations, and the use of alternative communication channels.

· **Multi-Source Technical Reserves:** Avoiding reliance on a single technology provider or nation, as such dependency creates a dangerous strategic vulnerability.

· **Pre-established Local Response Protocol:** Defining a clear decision-making hierarchy and specific authorities to activate pre-legislated cyber emergency laws; a crisis is not the time to draft protocols.

· **Societal Security Culture:** Implementing continuous awareness campaigns, staff training, and integrating security fundamentals into educational curricula so that citizen awareness becomes the first line of defense.

Post-Quantum Cryptography

A looming threat exists—one that is not “cyber” in the traditional sense: super-powerful quantum computers. Once matured, these machines will be capable of cracking most encryption algorithms currently used to protect government data. The terrifying danger is not merely in the future; it is unfolding right now through an adversarial strategy known as “Harvest Now, Decrypt Later.” Adversaries today are capturing and storing encrypted data transmitted via undersea cables and satellites, awaiting the future moment when encryption is broken by the advent of super-powerful computing systems. Consequently, an immediate plan is required, comprising the following steps:

1. Conducting a comprehensive inventory of the algorithms used in critical government systems.

2. Classifying data according to its “independent lifespan,” the duration for which the data remains sensitive and warrants protection.

3. Initiating a gradual transition to approved quantum-resistant standards and ensuring that any new development supports cryptographic resilience from day one.

Conclusion and Strategic Recommendation

The real threat today is not the sudden collapse of a state, but instead the loss of its strategic room for maneuver and the erosion of trust in its institutions at a critical juncture. The difference between a state that thrives in this environment and one that falters lies not in the volume of its resources, but in its management mindset: a state that treats cybersecurity and artificial intelligence merely as technical issues to be handled by a communications ministry or an IT department will find itself generations behind, whereas a state that treats them as matters of independence managed at the highest political level will position itself for readiness and influence. The window of opportunity is open now, yet a delay of just one year in this arena is equivalent to a decade-long lag in other fields, given the staggering pace of technological development and progress.