08\26

2 CBRNE DIARY

Dedicated to Global First Responders

August 2026

PART B

New vaccine for Burkholderia pseudomallei (melioidosis)

KSA goes nuclear

Pufferfish toxin antidote

Religious AI

Sudan CWAs

AI bacteriophage

Genetic bioweapons

An International CBRNE Institute publication

ICI International CBRNE INSTITUTE

CBRNE-Terrorism Newsletter

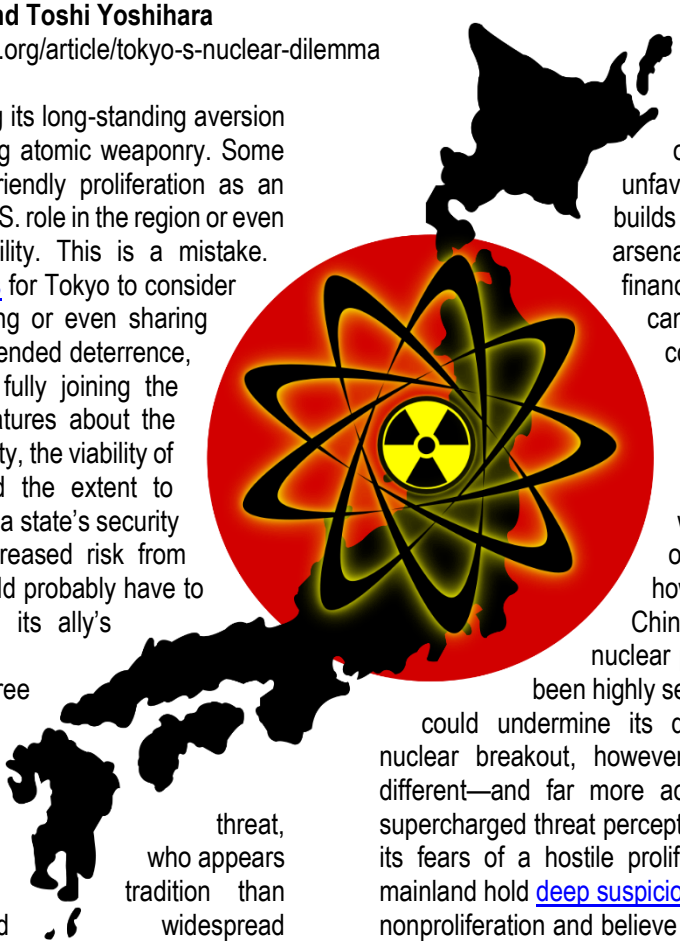
W O M D

DIRTY R-NEWS

C2BRNE DIARY

Tokyo's Nuclear Dilemma

By Evan Braden Montgomery and Toshi Yoshihara
Source: https://www.lawfaremedia.org/article/tokyo-s-nuclear-dilemma



July 26 – Japan may be rethinking its long-standing aversion to building, maintaining, or hosting atomic weaponry. Some American strategists see such friendly proliferation as an acceptable cost for reducing the U.S. role in the region or even as beneficial in preserving stability. This is a mistake. Although there are strong reasons for Tokyo to consider abandoning its aversion to hosting or even sharing nuclear weapons in service of extended deterrence, assessments of the benefits of fully joining the nuclear club often rest on caricatures about the risks of pursuing a nuclear capability, the viability of minimal nuclear deterrence, and the extent to which a nuclear arsenal can solve a state's security problems. Japan would face increased risk from China, and the United States would probably have to play a greater role in ensuring its ally's security.

Japan is surrounded by three nuclear-armed revisionist powers—China, Russia, and North Korea—that are each enhancing their arsenals. This threat, along with a new prime minister who appears less bound to foreign policy tradition than some of her predecessors and widespread questions regarding the U.S. role in the region and the world, may change Tokyo's approach to nuclear weapons. This is especially likely if South Korea opts to adjust its own nuclear status, leaving Japan as the only major power without a nuclear arsenal in an increasingly dangerous neighborhood.

Notably, over the past year, various reports have raised the possibility that Japan could rethink its nuclear status. In a November 2025 Diet session, Prime Minister Sanae Takaichi hinted at revisiting the language of the three non-nuclear principles ("not possessing, not producing, and not permitting the introduction of nuclear weapons" onto Japanese territory) as her government revised Japan's key national security documents. The following month, an unnamed adviser to the prime minister's office allegedly declared that Japan "should possess nuclear weapons" in response to a reporter's question. Takaichi later reiterated her commitment to the three non-nuclear principles, but just recently her defense minister advocated an open debate over the nuclear issue.

Should Japan pursue an independent nuclear weapons capability, this path-breaking choice would introduce three significant risks. First, it would open a window of opportunity for adversaries to attack before it crosses the nuclear threshold or fields a survivable arsenal. Second, if it crossed that threshold, and even after it established a "minimum deterrent," it would enter a valley of vulnerability given how unfavorably the nuclear options it builds would stack up against the arsenals of its neighbors. Third, if the financial costs of a nuclear buildup came at the expense of investing in conventional forces, which seems likely, Japan might find itself in a prolonged period of instability during which it would improve its protection from some threats while increasing its exposure to others. Consider, for instance, how the rivalry between Japan and China could evolve in response to a nuclear push by Tokyo. Beijing has long been highly sensitive to any developments that could undermine its deterrent posture. A Japanese nuclear breakout, however, would represent an entirely different—and far more acute—level of danger. Beijing's supercharged threat perceptions about Tokyo would magnify its fears of a hostile proliferator next door. Many on the mainland hold deep suspicions about Japan's commitment to nonproliferation and believe that it maintains a latent nuclear weapons capability. Such profound distrust of Tokyo likely reinforces Beijing's unwillingness to accept mutual vulnerability vis-a-vis Japan.

If Tokyo were to attempt a nuclear breakout, China could count on a range of coercive tools that would expose Japan to the three kinds of risks described above. Beijing could sanction Japanese companies, use subversion to divide Japan's body politic, or mass naval, air, and paramilitary forces near disputed territories, like the Senkaku Islands. Chinese leaders and commanders can mix and match these well-honed coercive options to disrupt the various stages of Tokyo's quest for the ultimate weapon before the window of opportunity slams shut. This pressure could erode Japan's will while the possibility of military preemption looms in the background. These coercive actions, moreover, would be taking place in China's nuclear shadow, the ultimate backstop to its strong-arming tactics. Indeed, Beijing could maneuver its nuclear triad in a demonstration of its resolve to intimidate Japanese leaders. China enjoys escalation dominance over Japan, meaning that its superior military prowess allows it to climb the escalation ladder higher than its rival. Beijing's ability to escalate further than Tokyo enables the Chinese side to threaten to impose unacceptable costs on the Japanese side,

bargain from a position of strength, and walk back its threats in exchange for Japan's concessions.

Even if Japan were to reach escape velocity in its nuclear breakout, Beijing's overwhelming nuclear superiority in numbers and quality would likely raise the bar, perhaps substantially, for what might constitute a satisfactory deterrent for Japanese planners. To achieve a level of sufficiency that meaningfully ensures mutual vulnerability, Tokyo might have to invest in a far larger, more diverse, and more survivable nuclear force structure and posture than presumed, which would prolong the period in which Japan would be vulnerable to China. And even if Tokyo were to field a credible second-strike capability, it would still not deter Chinese coercion that fell short of war, including demonstrations of force, economic reprisals, and political warfare. If Japan were to divert substantial resources to nuclear breakout at the expense of conventional military modernization and other tools for countering coercion, then it might remain vulnerable to those threats, which the nuclear shield is inadequate to deter. In other words, after crossing the nuclear threshold, Japanese leaders would face stark decisions and trade-offs as they balanced competing resource demands under persistent duress. Ultimately, China's imperative to preserve its nuclear superiority, its hypersensitivity to Japan as a potential nuclear rival, and its many tools that were unavailable to it just a generation ago suggest that Beijing will likely go to great lengths to maintain its preferred status quo. Meanwhile, Russia and North Korea might have fewer coercive options at their disposal, but that does not rule out either state taking steps to derail or counter Japan's nuclear ambitions—or the prospect that Tokyo's rivals could work together to keep the status quo intact. Of course, the United States could help to offset each of these dangers. That is, it could provide the resources, technology, and expertise necessary to support a nuclear buildup, not only to accelerate the process but also to avoid a situation in which Japan cannibalizes much-needed conventional defenses in pursuit of a viable nuclear capability. At the same time, the United States could double-down on its security commitment to get Japan safely across the nuclear threshold, backstop its new arsenal with U.S. nuclear forces to make it a better match for rivals, and make up for any gaps that might emerge in Tokyo's ability to address non-nuclear threats. Yet doing so would undercut the rationale that proliferation allows Washington to reduce its commitments, avoid being drawn deeper into regional competitions, and conserve its limited resources.

The potential blowback from Japanese proliferation uncovers a major paradox that advocates of friendly proliferation miss: Any potential benefits of nuclear proliferation for the United States would be highly dependent on Washington assuming the very costs and risks that friendly proliferation would be meant to offset. Debates over alternatives to extended deterrence must be clear-eyed about the hidden and not-so-hidden dangers of friendly proliferation.

Evan Braden Montgomery is vice president for research and studies at the Center for Strategic and Budgetary Assessments.
Toshi Yoshihara is a senior fellow at the Center for Strategic and Budgetary Assessments.

What would be the effect of a nuclear Saudi Arabia in the Middle East?

The Editor

A nuclear-armed Saudi Arabia would fundamentally alter the strategic balance of the Middle East, marking one of the most consequential shifts in regional security since the emergence of Israel's undeclared nuclear deterrent. Whether Saudi Arabia developed indigenous nuclear weapons or acquired them externally, the result would almost certainly be the emergence of a multipolar nuclear environment characterized by greater deterrence but also increased instability.

From Riyadh's perspective, a nuclear capability would primarily serve as a strategic deterrent against Iran. Saudi leaders have repeatedly indicated that if Iran were to acquire operational nuclear weapons, the Kingdom would seek a comparable capability. Such a development could restore a degree of strategic balance across the Gulf by reducing the perceived asymmetry between the two regional rivals. However, deterrence in the Middle East would differ from the relatively stable Cold War model because regional crises tend to be more frequent, involve multiple state and non-state actors, and are often influenced by ideological and sectarian dynamics.

The greatest consequence would likely be nuclear proliferation. Countries such as Egypt and Turkey might reconsider their long-standing non-nuclear policies, arguing that regional security requires equivalent capabilities. This could gradually transform the Middle East from a region with a single presumed nuclear power into one with several competing nuclear states, significantly increasing the complexity of crisis management and strategic signaling.

Relations with the United States would also evolve. While Washington has traditionally opposed nuclear proliferation among its allies, it might face difficult strategic choices if Saudi nuclearization were viewed as a response to an overt Iranian nuclear arsenal. At the same time, U.S. efforts to preserve the global non-proliferation regime would become

considerably more challenging, potentially weakening the credibility of the Treaty on the Non-Proliferation of Nuclear Weapons.

Israel would confront a substantially different security environment. Although Israel has maintained military superiority and strategic ambiguity for decades, the presence of another technologically advanced regional nuclear power would require significant adjustments in intelligence, missile defense, early warning systems, and strategic planning. While direct military confrontation might become less likely due to mutual deterrence, indirect competition through proxies and cyber operations could intensify.

The Gulf Cooperation Council would likely become more cohesive in terms of collective defense but also more dependent on sophisticated missile defense, early-warning networks, and nuclear command-and-control arrangements. International actors would increase efforts to establish regional confidence-building measures, communication hotlines, and arms-control mechanisms to reduce the risk of accidental escalation.

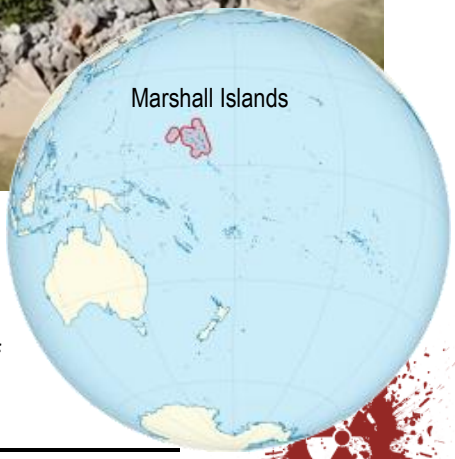
Economically, markets would initially react negatively because of increased geopolitical uncertainty, potentially affecting energy prices, investment, and insurance costs for maritime transport through the Strait of Hormuz and the Red Sea. Over the longer term, however, markets might stabilize if a credible deterrence relationship reduced the likelihood of large-scale conventional war.

The principal danger would not necessarily be deliberate nuclear war, which remains unlikely because of the catastrophic consequences for all parties. Rather, the greater risks would stem from miscalculation during conventional conflicts, cyberattacks affecting nuclear command-and-control systems, accidental escalation, or unauthorized use. These risks would be amplified by the region's dense concentration of critical energy infrastructure, short missile flight times, and multiple ongoing proxy conflicts.

In summary, a nuclear Saudi Arabia would probably increase strategic deterrence against Iran while simultaneously accelerating regional proliferation, complicating alliance structures, and raising the probability of nuclear crises. The Middle East would become more difficult to stabilize, with a greater emphasis on deterrence, crisis communication, missile defense, and arms-control initiatives to prevent escalation from conventional conflict to the nuclear threshold.

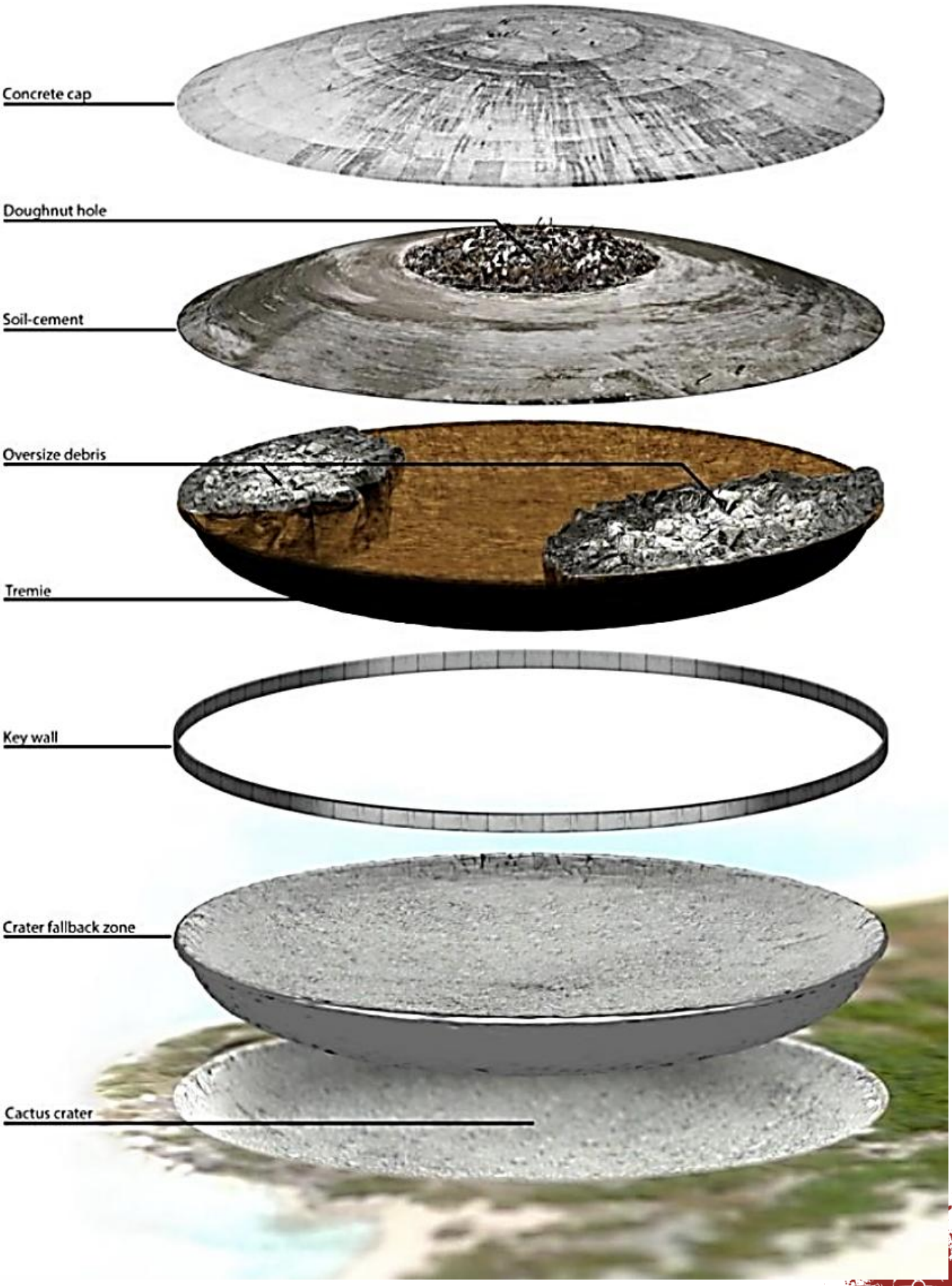
In 1977, the U.S. buried 120,000 tons of radioactive waste in a Pacific 'nuclear tomb.' Nearly 50 years later, the ocean is threatening to expose the Runit Dome

Source: https://en.as.com/latest_news/in-1977-the-us-buried-28-million-cubic-feet-of-radioactive-waste-in-a-pacific-nuclear-tomb-nearly-50-years-later-the-ocean-is-threatening-to-expose-the-runit-dome-f202607-n/



July 28 – Surrounded by turquoise waters and coral reefs in the remote Pacific Ocean sits a little-known concrete structure that **remains one of the world's most controversial nuclear waste sites.**

Known as "The Tomb" or the Runit Dome, the circular concrete cap measures **377 feet in diameter and is 18 inches thick**. For nearly half a century, it has attempted to contain one of the **most hazardous remnants of the Cold War.**


Concrete cap
Doughnut hole
Soil-cement
Oversize debris
Tremie
Key wall
Crater fallback zone
Cactus crater

To understand how the dome came to exist, it's necessary to go back to the 1940s and 1950s. During the nuclear arms race with the Soviet Union, the United States selected the Marshall Islands as its primary nuclear testing ground. In total, 67 nuclear tests were carried out in the region, **with 43 detonated at Enewetak Ato ll, where the Runit Dome is located.**
Among them was the devastating Castle Bravo test at neighboring Bikini Atoll, a hydrogen bomb roughly **1,000 times more powerful than the atomic bomb dropped on Hiroshima.**
The repeated tests vaporized entire islands, forced local communities from their homes, and **blanketed both the land and seafloor with radioactive fallout.**

A giant concrete patch
By the late 1970s, amid growing international pressure and demands from displaced islanders, the **U.S. government launched a cleanup operation.**
Military engineers removed contaminated topsoil, radioactive ash, vegetation and abandoned military debris from across the atoll.
The material was then dumped into the 328-foot-wide Cactus Crater, created by the 1958 Cactus nuclear test on Runit Island.
Finally, workers covered the crater with **358 massive concrete panels**, creating the dome that remains today.
But the project contained a critical flaw.
To reduce costs, engineers **never sealed the bottom of the crater**. Instead of installing an impermeable barrier, **they assumed the atoll's natural rock would prevent radioactive material from escaping.**
The assumption proved incorrect. Unlike solid granite, an atoll is built on **porous coral limestone**, allowing seawater to move freely through the ground.

Rising seas threaten the structure
Nearly five decades later, time and climate change have exposed the dome's weaknesses.
Scientists have identified cracks in the concrete cap, water infiltration through the coral beneath the crater and growing concerns over rising sea levels, **all of which threaten the long-term stability of the site.**
Buried beneath the structure are an estimated **2.8 million cubic feet of radioactive waste**, equivalent to roughly **120,000 tons** of contaminated material.The debris includes radioactive ash, contaminated soil, vegetation, military scrap and concrete rubble.
Experts are particularly concerned about plutonium-239, **an extremely toxic radioactive isotope with a half-life of more than 24,000 years.**

Scientists raise the alarm
The dome's deteriorating condition has increasingly alarmed researchers. "The dome was essentially a temporary fix that was never designed to last," marine physicist Ken Buesseler of the Woods Hole Oceanographic Institution has warned. "Seawater is literally flowing through the radioactive waste."

Michael Gerrard, director of Columbia University's Sabin Center for Climate Change Law, has **also questioned the dome's long-term viability.** "The bottom of the dome is simply the crater left by the explosion," Gerrard said. "There is no impermeable barrier underneath, so as sea levels rise, the Pacific Ocean is already washing through the inside of the tomb from below."

Dispute between the United States and the Marshall Islands

The deteriorating structure has become the center of an ongoing dispute between the **governments of the Marshall Islands and the United States.** Officials in the Marshall Islands have called for a long-term containment plan or for the radioactive waste to be removed entirely and transported to the U.S. mainland.

Washington, however, has long maintained that responsibility for maintaining the structure was transferred to the Marshall Islands **following the country's independence agreements in the 1980s.**

Today, the Runit Dome still sits silently in the middle of the Pacific.

It remains a concrete reminder of the nuclear age and of a lesson that continues to resonate decades later: humanity may be able to bury its most dangerous mistakes beneath thousands of tons of concrete, **but the ocean has a way of finding every crack.**

Climate change exacerbates nuclear risks in the Arctic. And we are not ready

By Nikita Gryazin, Iren Marinova

Source: https://thebulletin.org/2026/07/climate-change-exacerbates-nuclear-risks-in-the-arctic-and-we-are-not-ready/

July 29 – Climate change is transforming the Arctic: Melting sea ice, thawing permafrost, and changing maritime accessibility are opening new territories, resources, and shipping routes while reshaping the region's strategic landscape. These changes have placed the Arctic at the center of a geopolitical competition between Russia, NATO states, and increasingly China. But the significance of these developments extends beyond economic interests and regional rivalry.

The Arctic hosts critical military infrastructure, strategic nuclear assets, and key pathways for power projection, making it a geopolitical magnifying glass in the evolving global security environment. But while climate change is widely recognized as a threat multiplier in the Arctic, regional governance has not yet adapted to this new reality. Climate change, geopolitical competition, and nuclear risk are typically treated as separate policy challenges, managed by separate institutions with different mandates. The global climate governance institutions approach the Arctic primarily through a focus on climate change effects and environmental management, while the Treaty on the Non-Proliferation of Nuclear Weapons (NPT), the bedrock of the global nuclear order, addresses nuclear risk through the traditional pillars of non-proliferation, disarmament, and peaceful uses of nuclear energy. At the regional level, the Arctic Council, the leading intergovernmental forum for Arctic governance and cooperation, explicitly excludes military security from its mandate. The result is a growing governance gap that no longer reflects the magnitude and complexity of issues in the Arctic.

The governance of the Arctic remains organized around institutional silos that treat climate, security, and nuclear issues as separate rather than mutually reinforcing threats. This disconnect leaves policymakers ill-prepared to manage the security consequences of a rapidly changing Arctic as global tensions are rising and the nuclear order is already under strain. Addressing the governance deficiency in the Arctic is now both a regional and a nuclear risk-reduction imperative. It requires a complementary risk-reduction dialogue—potentially structured as a Track 1.5 mechanism—alongside Arctic-specific planning within NATO, greater attention to regional risks through the NPT process, and renewed bilateral risk-reduction channels.

The Arctic under a nuclear umbrella

A main source of increased nuclear dangers in the Arctic is the intensified geopolitical competition among interested nuclear-armed powers. Increased cooperation between Beijing and Moscow and the expansion of NATO to encompass seven of the eight Arctic states have reshaped Arctic geopolitics around growing competition between Russia and, increasingly, China on one side, and the United States and NATO on the other. This happens as the entire Arctic region now finds itself under a nuclear umbrella.

This strategic competition carries important nuclear dimensions.

The Arctic remains central to Russia's second-strike nuclear deterrent, hosts critical military infrastructure, including early-warning systems and two-thirds of its sea-based

nuclear forces. The region also serves as a key operating environment for Russia's strategic submarines, which have historically relied on Arctic ice cover for concealment from NATO anti-submarine warfare assets, using the region's unique acoustic environment and under-ice operating conditions to reduce detectability. Climate change has begun to erode the very geography underpinning this strategy. At the same time, China has expanded its economic, scientific, and military footprint in the region through closer cooperation with Russia and growing interest in Arctic shipping routes and resource development.

For the United States, the Arctic is central to homeland defense and nuclear deterrence, serving as a key theater for missile warning, strategic surveillance, and submarine operations. In response to Russia's growing Arctic posture, NATO has significantly expanded its own presence through exercises, deployments, and anti-submarine warfare activities. Finland and Sweden, by joining NATO in 2023 and 2024, respectively, have altered deterrence dynamics and the broader nuclear balance in the Arctic. Their accession has changed the Nordic region from being a peripheral space of nuclear restraint to becoming a strategically integrated frontline of the alliance's deterrence posture. For decades, both Nordic countries remained non-aligned and staunch promoters of non-proliferation, and even NATO members Denmark and Norway maintained cautious policies on the peacetime stationing of nuclear weapons on their territories. The extension of NATO's nuclear umbrella resulting from their accession has marked a critical strategic shift in the Arctic. While nuclear stationing in the region remains politically sensitive, NATO's expansion across the European Arctic increases the strategic significance of the High North and heightens the potential for escalation surrounding Russia's Arctic-based nuclear assets.

In early 2026, the US administration's coercive rhetoric over Greenland generated significant political tensions across the North Atlantic. But the Greenland crisis, combined with general shifts in US foreign policy towards Europe in the last year, has raised serious concerns among European allies over the long-term credibility of the US nuclear umbrella. The Nordic interest in complementary European deterrence arrangements involving France and the United Kingdom has also intensified as a result.

Nuclear dangers in the Arctic

A less predictable Arctic means a less predictable deterrence environment. Melting ice and increased maritime access raise the likelihood of submarine detection, close military encounters, and miscalculation between nuclear-armed powers operating in increasingly contested waters. These strategic changes in the Arctic extend beyond deterrence calculus and into an urgent need for regional cooperation.

Novaya Zemlya, the archipelago hosting Russia's nuclear test infrastructure, sits atop a far older problem. Between 1955 and 1990, the Soviet Union conducted 224 nuclear detonations there, releasing a total of around 265 megatons of explosive energy, and scuttled more than 100 decommissioned nuclear submarines in the surrounding Barents and Kara seas. Plutonium and cesium from both sources are detectable in regional sea sediments and soil beneath glaciers. Permafrost, which used to keep much of this contamination stable, is now thawing. Researchers showed that the Arctic's Cold War nuclear legacy is one of the least-understood consequences of global warming, as radioactive isotopes locked in frozen soil for decades are now finding new pathways in the environment. Novaya Zemlya remains largely closed to independent researchers, and Russia has offered minimal transparency about current contamination levels at the test site.

The Arctic is also increasingly becoming a theater for nuclear signaling. Russia has repeatedly invoked its nuclear capabilities in the context of confrontation with NATO, embracing more aggressive forms of signaling and appearing more willing to blur the thresholds surrounding potential nuclear use. The result is a region where climate disruption, military competition, and nuclear deterrence are becoming mutually reinforcing pressures.

A fragmented governance system frozen in place

Existing governance institutions have not yet adapted to the rapidly evolving nature of these multifaceted threats in the Arctic. The region's governance remains reliant on a fragmented and complex ecosystem of international treaties, UN bodies, and regional forums that primarily dictate shipping, resource management, and environmental protection. But the unique complexity of challenges in the Arctic and the rising geopolitical tensions require a more integrated approach.

Geopolitical tensions have removed many of the mechanisms that historically helped manage Arctic issues. Since Russia's full-scale invasion of Ukraine in 2022, the Arctic Council has been largely paralyzed, scientific cooperation has deteriorated, and global arms control structures have continued to erode. No treaty governs nuclear activities in the region, and the collapse of US-Russia arms control frameworks has left few institutional safeguards in place.

These overlapping gaps compound one another: Regional governance relies on multiple different institutional arrangements; the Arctic Council cannot discuss military activity; and the NPT review process does not address region-specific escalation risks. This results in a structural blind spot that leaves states without agreed rules of the road at precisely the moment when the risk of miscalculation is rising. The absence of any dedicated forum that brings together climate scientists, arms control experts, and defense officials means that the linkages between environmental change and nuclear risk remain invisible to policymakers until a crisis makes them undeniable.

Closing the governance gap

The disconnect between siloed governance bodies and self-reinforcing risks in the Arctic demands concrete responses across several institutional levels. Despite its current constraints, the Arctic Council remains the region's most successful institutional forum for circumpolar cooperation. Institutional bridges should be built around it as a cornerstone. But cooperation through the Arctic Council has been resilient precisely because its mandate explicitly excludes military security and instead focuses on scientific cooperation, sustainable development, and participation of indigenous populations. Rather than expanding the Arctic Council's mandate to encompass military security, the Arctic states should instead focus on closing the governance gap between climate change, geopolitical competition, and nuclear risks. This could be achieved through a complementary Arctic risk-reduction dialogue, potentially structured as a Track 1.5 mechanism, that operates alongside the Arctic Council and brings together foreign ministers, defense officials, coast guards, scientific experts, and representatives of indigenous communities. It could potentially be convened by an Arctic state with a strong record of regional diplomacy, such as Norway or Iceland, or a UN-affiliated body. The purpose would be to bring together experts in different areas to discuss the changing and multi-faceted threat environment in the Arctic.

Initially, such a forum could focus on practical areas of shared concern, including maritime incidents, search and rescue, environmental problems, critical infrastructure, and nuclear risk mitigation. Creating formal channels of communication would provide a venue for identifying and managing the increasingly interconnected climate, geopolitical, and nuclear threats emerging in the Arctic. Complementary institutional responses should be established above the immediate regional level as well. Building on the Strategic Concept's recognition of climate change as a threat multiplier, NATO should translate this commitment into Arctic-specific planning by incorporating climate-driven deterrence instability into its strategic planning and pursue dedicated Arctic incident-prevention protocols, including direct communication channels modelled on existing incidents-at-sea agreements, to reduce the risk of miscalculation. At the multilateral level, the NPT review process should give greater attention to the interconnected nuclear risks in the Arctic by encouraging voluntary transparency, expert assessment, and reporting on climate-related vulnerabilities, radioactive contamination and nuclear legacies in the region, including at sites such as Novaya Zemlya. Finally, even amid the current political environment, the United States and Russia should prioritize the Arctic by pursuing or renewing Arctic risk-reduction channels, including incidents-at-sea arrangements and joint environmental monitoring of thawing contaminated sites. While these are minimal confidence-building measures that do not require broader arms control consensus, they could meaningfully reduce the risk of miscalculation and escalation. Rapid warming in the Arctic has opened the door to intense geopolitical competition in the High North at the same time that the region is being placed under a broader nuclear umbrella. In the Arctic, these issues can no longer be addressed as separate concerns but rather as reinforcing one another and heightening regional and global insecurity. Existing governance institutions in the region must adapt to these new security challenges.

Nikita Gryazin is a policy fellow at the European Leadership Network (ELN).
Iren Marinova is a Deutsche Akademischer Austauschdienst (DAAD) Postdoctoral Fellow at the Henry A. Kissinger Center for Global Affairs at the Johns Hopkins University School of Advanced International Studies (SAIS) in Washington, DC.

Balancing global uranium enrichment needs and proliferation risks: A framework

By Logan Mintz and Matthew Sharp

Source: https://thebulletin.org/2026/07/balancing-global-uranium-enrichment-needs-and-proliferation-risks-a-framework

July 27 – The Trump administration is pursuing two contradictory policies with uranium enrichment: It has insisted that zero uranium enrichment in Iran is necessary to achieve its goal of preventing Iranian access to a nuclear weapon. But the administration has also shown an apparent tolerance for the spread of the same uranium enrichment technology elsewhere as part of its push for expanded nuclear exports. There is an essential tension between the administration's readiness to use military force to prevent enrichment in Iran while condoning the technology's spread elsewhere. President Donald Trump showed greater comfort with the spread of enrichment in two high-level meetings in late 2025. Following his October 29 meeting with South Korean President Lee Jae Myung, the White House stated that the United States now supports "the process that will lead to the Republic of Korea's civil uranium enrichment and spent fuel reprocessing." This was a notable departure from past US policy against Republic of Korea enrichment.

A month later, during Saudi Crown Prince Mohammed bin Salman's visit to Washington, US and Saudi officials previewed a pending nuclear cooperation agreement between

the two countries, also known as a "123 agreement." Past administrations have been unable to conclude such an agreement with Saudi Arabia in part due to Riyadh's insistence on the need for domestic enrichment, which US policy opposed. While no text has yet been made public officially, President Trump recently approved the agreement, and there are indications that the nuclear cooperation deal will, at the least, leave open the possibility of future Saudi enrichment.

Gas centrifuges for uranium enrichment seized from a cargo ship en route to Libya, in 2003. The illicit cargo was traced directly to the A. Q. Khan proliferation network—a global black-market network led by Abdul Qadeer Khan, the father of Pakistan's nuclear weapons program. The designs and technical blueprints of the components originated from Khan Research Laboratories in Pakistan, based on centrifuge designs originally derived from European enrichment technology. (Credit: US Energy Department)

Consistent with a shift in the administration's direction on enrichment is a May 2025 executive order directing aggressive negotiation of at least 19 additional 123 agreements by the end of 2028—a significant increase from the 26 such agreements currently in force. These deals could be a boon for US leadership in the responsible global expansion of nuclear energy and potentially help strengthen nonproliferation standards. However, with the clock ticking to meet the president's tight timeline, US negotiators may seek to relax past positions on enrichment if they are perceived as impeding agreement.

The trouble with spreading enrichment

Enrichment is a truly dual-use technology: The same process that can turn raw uranium into fuel for a power plant (if uranium is enriched to 3 to 5 percent uranium 235) can also be used to make material for a bomb (if enriched to 90 percent or higher).

A state with civilian enrichment capabilities has several otherwise unavailable pathways through which to pursue nuclear weapons: It could divert material, misuse facilities within declared civil programs, or establish a covert enrichment facility to support a military program. Enrichment technology exported to a US ally could be proliferated to a third country, as was the case with the

A.Q. Khan network's illicit transfer of European enrichment technology to Pakistan. Such scenarios could unfold rapidly, providing short timeframes for detection.

Once acquired, states are loath to relinquish a capability considered prestigious, and knowledge of enrichment technology cannot be erased. There are also norms to consider: Today's allies with enrichment technology could become tomorrow's adversaries—and today's adversaries could take advantage of a precedent once set. For these reasons, US practice has been to stave off future proliferation challenges by preventing any new enrichment entrants, even friendly ones, in the first place.

Why states pursue enrichment

Countries commonly frame domestic enrichment as a right to the peaceful uses of nuclear energy. Economically, states want to have their share of the multi-billion-dollar enrichment industry. In terms of energy security, countries with a nuclear power program are concerned that dependence on another state for enrichment puts them at risk of a supply disruption. According to this line of thinking, domestic control of the entire fuel cycle—not just enrichment—is the only surefire answer to this concern.

A third argument, often unstated, is that enrichment enhances national security by also providing the technical ability to produce material for a nuclear weapon. While seeking access to enrichment, South Korean and Saudi officials have each entertained the idea of having their own nuclear weapons program, and experts speculate that Iran's desire to linger on the nuclear threshold motivated its enrichment program.

These motivations should be taken seriously, but without overstatement. In almost every case, it is cheaper to purchase reactor fuel from foreign suppliers than to produce it domestically. Enrichment facilities cost billions of dollars, take years to build, and have long, uncertain return horizons. Energy supply arrangements have proven extremely reliable, even years after Russia began its invasion of Ukraine. A state seeking security benefits from enrichment should also consider whether attacks like those by Israel and the United States on Iran's enrichment facilities suggest this hedge may leave a state less safe, not safer.

Limiting enrichment through negotiation

Any global expansion in nuclear energy will require enriched uranium to fuel reactors. The question, therefore, is not whether but where that uranium enrichment will take place. The United States, which is both a competitive nuclear supplier and a long-time champion of nonproliferation, has several tools to navigate rising international demand for enrichment technology.

Early US legal frameworks recognized the sensitivity of enrichment, protecting enrichment-related information in the same way that it protects nuclear weapons design information—and demanding that partners seek advance US consent before enriching US material. The United States has historically discouraged the development of new enrichment programs by neither providing advance consent to enrich US material nor transferring enrichment technology to countries that did not already have that capability.

But the US nuclear industry is less globally dominant than it was when these policies were initiated, meaning that a US partner could forego the ability to enrich US material while separately partnering with another state or working to indigenously enrich non-US material. The United States has therefore sought commitments from partners against acquiring enrichment from other states or developing it themselves. Some worry this position could turn off potential partners and open the door to more permissive suppliers like Russia or China. Indeed, few partners have agreed to the more stringent approach—the United Arab Emirates being one of the few examples.

Such fears of ceding ground to China or Russia through a stronger negotiating stance could be overblown. The multilateral Nuclear Suppliers Group, which includes China and Russia, has tightened controls on enrichment and shown little appetite for its transfer.

Neither Russia nor China uses indigenous enrichment as an enticement for potential nuclear energy customers. Instead, Russia includes both front- and back-end fuel cycle services in its long-term contracts, and China offers front-end services. Rebuilding US enrichment capability, if coupled with similar front-end guarantees for partners, could help obviate the need for domestic enrichment by partners.

The United States must use all available leverage during negotiations to ensure the benefits of a US civilian nuclear agreement to a partner state outweigh its nonproliferation requirements. Quality technology and competitive financing are necessary elements, but potential partners may seek more. The United States can seek to reduce the allure of an enrichment program by offering instead a suite of political, economic, and military sweeteners—even outside of the official nuclear energy package—that other suppliers may not be able to provide.

Security incentives could be addressed through US conventional military technology; while often treated separately, potential nuclear customers are often also defense cooperation partners. Security considerations may be even better dealt with through security guarantees or non-aggression pacts. Cooperation on data centers and advanced computing technology, potentially powered by nuclear energy, could also entice states with economic motivations.

Viewing international cooperation in civilian nuclear energy technology in a vacuum undersells US bargaining power and ignores the broader geopolitical context of such cooperation. In addition to improving US bargaining power through nuclear technology exports, Washington should consider pairing nuclear exports with the non-

nuclear offerings they are intended to power—such as data centers or hydrogen production plants—advancing US nonproliferation goals without giving up the sale.

Accommodating demand for enrichment

The United States can successfully hold a hard line on transferring enrichment technology by working with partners to develop alternative arrangements that address the underlying economic, energy security, and national security factors driving interest in enrichment.

Alternative fuel supply options can directly address energy security concerns, while also financially disincentivizing enrichment aspirants by creating well-funded competition. In 2023, the United States, the United Kingdom, France, Canada, and Japan—collectively known as the "Sapporo 5"—formed a coalition of uranium producers to secure reliable fuel supply chains and diversify from Russian supply. The group has secured $5.6 billion in government and private investment for new enrichment and conversion capacity. The US element of this effort seeks to leverage significant funding from the US Energy Department to generate more domestic and international investments in US companies.

Instead of pursuing domestic enrichment, international partners should take advantage of this opportunity, acquiring an equity stake in existing US enrichment companies. Direct investment by states without enrichment in existing companies would enhance confidence in access to the product, without spreading the proliferation-sensitive technology, addressing both fuel supply and economic considerations without domestic enrichment. Less ambitiously, countries could instead seek contractual arrangements with existing enrichment companies to build confidence that their supply would not be disrupted; such contracts could provide, for example, improved uranium distribution mechanisms or more robust contingency planning, including arrangements for a secondary supplier to complete a hypothetical disrupted transaction.

Regional fuel cycles could also reduce energy security concerns by introducing mutual dependence between states across the entire fuel cycle, including enrichment, uranium mining, and fuel fabrication. For example, an arrangement could link Japanese enrichment, South Korean fuel fabrication, Mongolian uranium mining and milling, and Taiwanese uranium conversion. Experts have also suggested similar arrangements for states in the Middle East. Each partner state would have access to the products—but not the technology—of the entire fuel cycle, leaving it dependent on the others for the fuel cycle steps it lacks while allowing it to benefit economically from its own link in the chain.

Some have suggested that partner demand for guaranteed access to enrichment in the country could be squared with technology protection needs through what are known as "black-box facilities," in which an extant supplier builds and operates the facility, and the host country owns the land and the product but has no access to the technology. Such proposals for enrichment are untested and therefore raise new regulatory, security, and legal hurdles.

International control of parts of the fuel cycle would go further. In 2019, the International Atomic Energy Agency (IAEA) opened a bank containing 90 metric tons of uranium enriched up to 5 percent available to member states experiencing reactor fuel supply disruption. This bank could be exhausted by just one fuel reload of a large reactor, but more expansive arrangements could provide more assurance. Contingency supply reduces the need for domestic enrichment for energy security purposes by preventing a supplier from using reactor fuel delivery as leverage against a state that needs it. Expanding such fuel banks, or even international enrichment, could help address energy security concerns.

Adapting with enhanced monitoring and verification

If enrichment technology does spread, the international community should enhance existing nuclear verification efforts and introduce new monitoring tools to quickly detect any attempt to produce nuclear weapons. Because states with enrichment capabilities can acquire weapons-usable material more rapidly than those without, they require more intense scrutiny.

A starting point is broader implementation of existing tools like the Additional Protocol, even making it a global condition of nuclear exports. This protocol enhances IAEA measures for detection of undeclared nuclear activities. If the United States stops conditioning deals on adoption of the Additional Protocol (apparently likely in the upcoming US-Saudi 123 agreement), expanded access to enrichment would only become riskier.

Iran's example shows the Additional Protocol is necessary but insufficient. The Joint Comprehensive Plan of Action (JCPOA), also known as the Iran nuclear deal, allowed Iranian enrichment up to 3.67 percent. But the deal also instituted a verification regime that went well beyond the Additional Protocol in its monitoring of centrifuge manufacturing, enrichment, and uranium mining and processing. More widespread enrichment would demand broader application of such verification measures. Other relevant upgrades to the Additional Protocol could include wide-area environmental sampling, weaponization-related inspection authorities, and updated reporting of nuclear-related imports, with particular importance in states where enrichment takes place.

Broader access to enrichment and, therefore, condensed proliferation timelines should motivate the development of new tools that address the issue of nuclear intent.

The current nonproliferation regime is agnostic to intent, benefitting from the fact that most of the international community is far from a nuclear weapons capability: As long as monitoring can confirm that nuclear materials stay where they should, there is no need—and the IAEA has no mandate—to confirm that a state's

nuclear activities are consistent or inconsistent with a peaceful purpose. With broader access to enrichment, states may need to develop new mechanisms to demonstrate consistency between their capabilities and their intentions and, therefore, that their activities are more consistent with civilian purposes than military ones.

States with existing enrichment capabilities should set the standard for newcomers. Establishing a monitoring precedent in these countries, beneficial on its own merits, would lay the groundwork for early warning if new states acquire enrichment technology. Tripling nuclear energy by 2050—as more than 30 countries have pledged to do—will strain already-stretched IAEA resources; new enrichment facilities on top of numerous new reactors would dramatically worsen the burden, as the IAEA spends up to 15 times more inspection days safeguarding an enrichment facility than a reactor.

As the Acheson-Lilienthal report posited 80 years ago, no amount of verification can fully compensate for the increased risk of broadly available enrichment capabilities. North Korea, by ejecting IAEA inspectors and then pursuing nuclear weapons, showed that even strong verification measures cannot prevent a country determined to join the nuclear weapons club. Allowing the spread of enrichment technology only intensifies this reality; stronger verification tools are necessary, but not sufficient to address this challenge.

The significant challenges posed by uranium enrichment do not justify opposing the expansion of nuclear energy in and of themselves. Rather, they simply call for not expanding nuclear energy at the cost of nonproliferation. The current wave of renewed enthusiasm for nuclear energy provides an opportunity to rally support for bolder action on nonproliferation than could have been considered only a few years ago.

Recent US military operations in Iran demonstrate the seriousness of a breakout enrichment capability; a nuclear renaissance should not be undertaken in a way that creates more such crises. Ambitious plans for small modular reactors that can power data servers or space stations will not succeed if they create their own proliferation crises. The reality of those stakes aligns US international trade and national security interests around blocking proliferation pathways, and it should prompt a broader conversation about how sustained investment in nuclear energy can be coupled with the means to harness the atom responsibly. Developing new tools of nuclear bargaining, fuel arrangements, and verification should be part of that conversation.

Logan Mintz is a director with the Nuclear Materials Security program at the Nuclear Threat Initiative (NTI).
Matthew Sharp is a senior nuclear fellow with the Security Studies Program at MIT's Center for Nuclear Security Policy (CNSP).

A film about nuclear war "too horrifying for the medium of broadcasting"

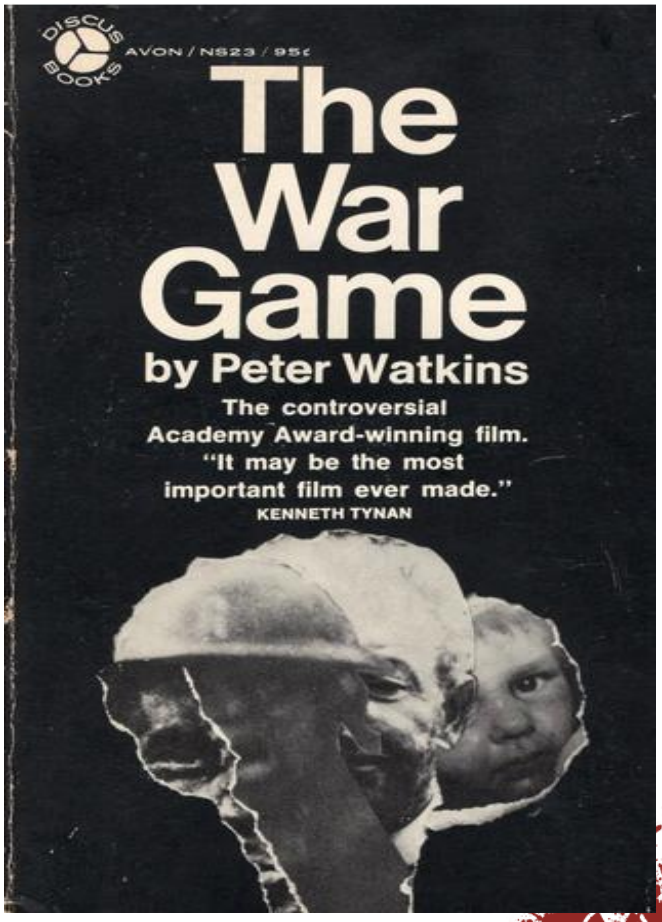
By Rebecca Fons

Source: https://thebulletin.org/premium/2026-07/a-film-about-nuclear-war-too-horrifying-for-the-medium-of-broadcasting

Oh, where are you coming from, soldier, gaunt soldier,
With weapons beyond any reach of my mind,
With weapons so deadly the world must grow older
And die in its tracks, if it does not turn kind?
—Song for Three Soldiers, Stephen Vincent Benét

July 08 – The 1990s were foundational movie-watching years for me. I was in my teens, and the films I consumed filled out my world, which at the time was narrow. I grew up in the small town of Winterset, Iowa (pop. 5,000) and it was through cinema that I first saw the gritty streets of New York City; witnessed a passionate kiss between lovers; heard Arabic, French, and Japanese; and experienced what has become my favorite activity: being told a story on a big screen in a room full of strangers.

That decade was a golden age for disaster movies: *Twister* (1996, man vs. nature), *Independence Day* (1996, man vs. alien), and *Armageddon* (1998, man vs. meteor). I logged



countless hours in movie theaters watching beautiful people scream, run, drive, scream, shoot, cleverly quip despite certain death, kiss, scream, and always (even if a few characters had shuffled off the mortal movie coil gracefully and cinematically along the way) the heroes triumph against The Big Bad Thing that has come to destroy them—be it tornado, extraterrestrial, or killer rock from space.

Watching these films was a thrilling experience: cramming into the back row of a theater with my nerdy friends, big, buttery popcorn and enormous Diet Coke in hand as we screeched in terror, clutching one another tightly while the ubiquitous James Horner or John Williams score soared, blubbering when Bruce Willis sacrificed himself for the greater good, and holding our breath, praying that our favorite characters would make it out alive. It was exhilarating. These films were glorious and glossy products that Hollywood reliably churned out and that we would happily watch over and over again.

As I've gotten older, my palette has matured. I gravitate toward arguably more intellectual fare, but out of a sense of nostalgia—or maybe duty—I buy tickets and go to the sequels, prequels and remakes of the films of my youth; showing up to the theater for each reboot, franchise installment, and sequel. And each time, I sit in my seat, decidedly disconnected. I do not clutch the arm of my husband in the seat next to me, I do not hold my breath and pray. These new films are glossier than ever, but something has changed for me. Maybe the scripts have gotten worse, maybe there have just been one too many sequels.

Or perhaps it's a third thing. Perhaps it's that I don't particularly want to pay $20 to see the screaming, the running, the driving, the shooting, the clever quipping. I neither seek out nor shy away from escapism in my cinema (I'm a film curator, my profession requires an open mind. But when footage of bombs dropping, genocide, and the loss of life is livestreamed every day on the small screen in our pockets, I realize I have become tired—exhausted, in fact—by the endless parade of CGI disasters released every Friday at a movie theater near you.

What's more, my standard of cinematic terror has been so raised by **Peter Watkins'** 1965 pseudo-documentary ***The War Game*** that every other big screen cataclysmic storyline feels like folly, other disaster movies come across as distraction.

Watkins, a Brit, was commissioned by the BBC to make *The War Game*. Most people knew nothing about the arms race, and Watkins planned to make an informational film about the effect of a nuclear strike on Britain. Boy, did he.

The finished film—a 47-minute, uncompromising nightmare—was immediately suppressed by the BBC, who refused to release it, declaring it "too horrifying for the medium of broadcasting." (The National Archives 1965). Though they denied it, saying the decision to pull *The War Game* from television was their own, there is evidence that the BBC screened the film for members of the British Ministry of Defence, who pressured them to ban the film. *The War Game*was eventually given limited release in cinemas, won an Academy Award for Best Documentary Feature, and finally aired on the BBC in 1985, twenty years later.

Watkins was staunchly anti-war, and throughout his career he used his medium to faithfully communicate his message. *Culloden*, the feature he made prior to *The War Game*, portrayed the brutality and futility of the battlefield; and with *The War Game*, he set out to detail the ultimate disaster: man vs. nuclear catastrophe; man vs. extinction. At a steady clip, Watkins, using a cast of non-actors, matter-of-factly dramatizes September 18 to Christmas Day—the 99 days of terror that follow the dropping of a Soviet nuclear bomb on the outskirts of Kent, England.

Shot in the style of a dry news report, Watkins steadies his camera dispassionately at his subjects, forcing the audience to bear witness as a narrator, without an ounce of sentiment or dramatic soundtrack, explains how little time the citizens of Kent have to take shelter (approximately three minutes, maybe fewer) at the drop of the bomb; how—because he was struck by the bomb's thermal radiation—this little boy's eyeballs melted; how 12 seconds later this building collapses due to the subsequent shockwave; or how 10 miles away in Rochester, an airburst has caused the firestorm you see here. And here we see, horrified viewers, Britain is striking back—dropping a bomb on the Soviets. Nuclear war has begun.

The flash from a bomb detonation blinds a civil defense worker and a passerby, in this still from "The War Game." Via BBC.

As the medical professionals interviewed in the film tell us: those who live, who make it past the initial blast, the chaos,

and the overwhelmed emergency rooms (for every 800 injured, there is one doctor) will succumb to their wounds, to radiation poisoning, or simply to the mental toll of what they have seen. Those who endure beyond these atrocities live to face the awful truth: After the bomb drops, there are no heroic gestures—this isn't that kind of disaster movie. No single man or woman rises up from the ashes to lead the weak to victory against armed men guarding the caches of food—the hungry are shot on the spot. A sick child does not miraculously overcome his ailments thanks to the power of love—no, the narrator impassively tells us that he has seven bedridden years to come, and then will die from an unknown illness, likely caused by highly enriched uranium.

Throughout the film Watkins weaves in "man on the street" interviews with the people of Kent, those not involved in the dramatization. The narrator asks them what they know about nuclear war, or the threat of nuclear-armed states. "Very little." "I'm afraid I don't know much about atomic radiation at all." "There won't be a war, no." A resident is seen showing off his bomb shelter—a shallow hole in the ground, covered by sandbags. You can imagine the film concluding in a 1965 screening room, the lights turning on to reveal BBC executives and UK government officials with ashen faces, all immediately agreeing that it would never see the light of day, if they had anything to say about it. You might even agree with them, to a degree: It is a hard film to watch. *The Guardian*'s Peter Bradshaw said that after he saw *The War Game*, "It seemed as if I had entered a new era of disillusioned adulthood" (Bradshaw 2025). You can choose, then, to not watch it. Like the *Choose Your Own Adventure* books of my childhood, you too can choose your own movie disaster when you go to the cinema. Or sit on the couch or pick up your phone: man vs. fill in the blank. As I write this, we're a month into another war with Iran, which the White House and its newly retitled Department of War has named "Operation Epic Fury." White House spokespeople are struggling to define the goal of the operation, or the action itself: It's a war, or definitely not a war; the goal is to disarm Iran's nuclear capabilities; or Iran has no nuclear capabilities because we already obliterated those.

Since the operation began, the White House has been posting stylized videos on its TikTok account to promote the war. In one, Nintendo Wii games are intercut with clips of "unclassified" infrared footage of bombs dropping on Iran military sites; a Wii golf game declares "Hole in one!" before cutting to the annihilation of a jet; a bowling ball hits all the pins, followed by grainy black-and-white footage of an artillery site ablaze: "Strike!" The video that has gotten the most attention is one featuring clips of Hollywood blockbusters. Titled "Justice the American Way," followed by the fire and American flag emojis, and set to music from the video game Halo, the TikTok begins with Robert Downey Jr. as *Ironman*'s Tony Stark, at a bay of computers. "Wake up, Daddy's home," he announces. In quick succession we see: Russell Crowe in *Gladiator* ("Strength and honor"), Mel Gibson in *Braveheart* ("What will you do without freedom?"), Tom Cruise in *Top Gun: Maverick* ("Maverick's inbound") before the screen fills with infrared footage of an enormous explosion. Scenes from *John Wick*, *Superman* ("I'm here to fight for truth, justice, and the American way."), *Star Wars,* and *Transformers* follow, alongside half-a-dozen shots of bombs dropping. It is not possible for viewers to know if in the strike footage there is loss of life, or to know what is left when the dust clears, but there is no question that they will recognize Mel Gibson, and everyone certainly knows that is Tom Cruise.

The United States military has a long, healthy relationship with Hollywood. Newsreels from wartime, movies romanticizing war; the military-entertainment complex is strong.

But these TikTok videos…there is something especially dangerous happening here. What associations do the average person have with *Top Gun: Maverick*? Big, buttery popcorn, Diet Coke, a sweeping Hans Zimmer score, Alpha masculinity. By connecting scenes of Capt. Pete "Maverick" Mitchell to shots of bomb impacts in Iran, a message is being sent loud and clear: Real war is just like the movies. Dropping bombs makes you feel like a Hollywood movie star, like an Avenger, like Tom Cruise.

With *The War Game*, Watkins' message was clear: When a bomb drops, the destruction is so complete, there can be no heroes. When you see the horrors that come from the pushing of the button, do you still think the man who pushes the button is brave? Sixty years later, the official White House TikTok account is publishing videos of military footage set to Los Del Río's *The Macarena*, and the audience loves it: Comments on these videos are rapturous in their praise. "Whoever runs this account deserves a raise and is a legend." "Iconic."

I am nostalgic about those silly disaster movies of my youth and the thrill they gave me—who doesn't like being entertained? Cinema has been thrilling audiences since 1896, when brothers Louis and Auguste Lumière showed an audience footage of a train arriving at a station and they fled, thinking the train was coming straight for them. When they realized there was no danger, they returned to their seats and an industry was born. *The War Game* was dangerous because it told the truth about the horrors of nuclear war; these TikTok videos are dangerous because they tell a lie—that war is a game and it's as easy as getting a hole in one.

Editor's note: *Like many of the older films that are mentioned in this article, "*The War Game*" can be seen by doing a search for it on YouTube. More recent movies can be viewed by paying the usual subscription services online, buying a DVD, or renting via the local library—or even [shocking] going in-person to a revival at the local cinema. And if one looks hard enough, there are still a few summertime drive-in movie theaters around, as the July issue of the* Bulletin *goes to press.*

References
Bradshaw, P. 2025. "Peter Watkins: an English film-making revolutionary from a tradition of uncompromising radicalism." *The Guardian*. https://www.theguardian.com/film/2025/nov/03/peter-watkins-film-making-revolutionary-war-game-culloden-punishment-park
The National Archives. 1965. "BBC film censored?" Sixties Britain, transcript of Parliamentary question asked in the House of Commons by William Hamilton MP to the Secretary of State for the Home Department. December 2. *The National Archives.* https://www.nationalarchives.gov.uk/education/resources/sixties-britain/bbc-film-censored/

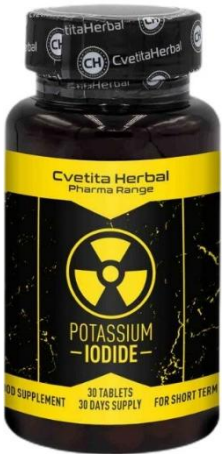
Rebecca Fons is a film curator, arts leader, and educator, now serving as Head of Cinema at the Barbican Centre in London, England, and as programming director for the historic Iowa Theater in her hometown of Winterset, IA. Fons is co-founder of the event series Destroy Your Art, and in 2022 was named Chicagoan of the Year in Film by *The Chicago Tribune*. She holds an MA from Columbia College, Chicago.

Hiroshima August 6, 1945: When "Little Boy" killed 80,000 people!

NRC to End Funding for Potassium Iodine Tablet Stockpile Distribution by FY 2028

Source: https://www.morganlewis.com/blogs/upandatom/2026/08/nrc-to-end-funding-for-potassium-iodine-tablet-stockpile-distribution-by-fy-2028

Aug 05 – The NRC has approved a staff recommendation to discontinue funding for potassium iodine (KI) stockpiles near nuclear power plants beginning in FY 2028. The NRC staff will pursue multiple options to ensure the NRC meets its delegated responsibility for KI tablet distribution under Section 127 of the 2002 Public Health Security and Bioterrorism Preparedness and Response Act (the Bioterrorism Act). This policy update is intended to reflect modern scientific evidence on the effectiveness of KI use following reactor release accidents and reduce inefficiencies within the KI tablet procurement process. This blog post analyzes the background, rationale, and implications of this policy shift.



Background

Potassium iodide (KI) is a stable iodine salt that, when taken in appropriate amounts at the right time, blocks the thyroid gland's uptake of radioactive iodine. Radioactive iodine is a byproduct that occurs during the fission of uranium atoms and could potentially be released into the environment during an accident at nuclear power plant facilities. Contrary to popular misconceptions, KI tablets only protect against

absorption of radioactive iodine by the thyroid, and do not provide protection against other types of radioactive materials. In 2001, the NRC amended its emergency preparedness regulations, 10 CFR 50.47(b)(10), to require that emergency plans for populations within the 10-mile emergency planning zones (EPZs) surrounding nuclear power plants consider KI tablets as a supplemental protective measure. KI tablet usage is intended to be a supplemental action to other protective actions like evacuation and sheltering. As a part of the adoption of this final rule, the NRC offered to fund states' initial purchases of KI tablets. This was noted as an exception to the long-standing policy that funding for state and local emergency planning is the responsibility of state and local governments. In 2002, federal KI policy developed further following the passage of the Bioterrorism Act. Section 127(a) directed the President to "make available to State and local governments potassium iodide tablets for stockpiling and distribution . . . within 20 miles of a nuclear power plant." The responsibility for implementing Sections 127(a) through 127(e) was delegated to the NRC in 2007. In a 2008 decision, the Office of Science and Technology Policy (OSTP) waived the requirements of Sections 127(a) and (d) thus declining to extend the program to a 20-mile radius. OSTP found that more effective measures than KI tablets exist beyond the 10-mile EPZ of a nuclear power plant to avoid exposures. Federal KI policy has not significantly evolved since.

NRC Policy Change: Ending Funding for Initial and Replenishment Stockpiles

Last month, the Commission approved the staff's recommendation to discontinue funding for both initial and replenishment stockpiles of KI tablets beginning in FY 2028. This change ends the decades-long exception to the NRC's policy that emergency planning costs are a state and local responsibility. Chairman Nieh's comments, echoing the staff's recommendation, stated that his decision hinged on two supporting principles. First, modern scientific insights on KI use and effectiveness suggest extremely low risk of thyroid cancer following radioiodine release, even in the absence of administering KI. Second, the current arrangement of NRC funding for KI stockpiles is insufficiently administered to ensure the efficient use of federal funds.

Scientific Evidence and Effectiveness of KI Tablets

In the decades following the adoption of NRC KI funding, further research has been conducted into the effectiveness of KI administration following the radiological release of radioiodine. Scientific evidence supports that timely KI administration, ideally before or within two hours of exposure to radioiodine, can reduce thyroid uptake of radioactive iodine. Recent research examining responses to modern nuclear accidents like Fukushima as well as the introduction of new, smaller reactors has led to the conclusion that while KI distribution methods can alter projected thyroid doses, real-world benefits are likely lower than modeled estimates. Given this research, NRC and international guidance continue to prioritize evacuation and avoidance of contaminated food over broad KI stockpiling.

Reducing Procurement Inefficiencies

The NRC staff also identified inefficiencies in the existing procurement process, noting that the agency's transfer of funds to the Department of Health and Human Services (HHS) for KI purchases introduces delays and complicates stockpile management. Additionally, there is no standardized process for states to request KI or estimate their actual needs, making budget planning unpredictable and undermining accountability for federal funds. Discontinuing NRC funding may, in interested communities, encourage more resilient, state-driven stockpiling and distribution plans and reduce the risk of fraud, waste, and abuse.

Transition Plan: Meeting the NRC's Delegated Bioterrorism Act Responsibility

The Commission directed the staff to pursue, listed below in priority order, several options to ensure that the NRC's Section 127 delegated authority is met:

1. Establishing an agreement with HHS or another federal agency to provide sufficient quantities of KI tablets to state and local governments without requiring requests under the Bioterrorism Act;
2. Developing and offering a technical basis and recommendation to the OSTP for a waiver of KI requirements under Sections 127(a) and 127(d) of the Bioterrorism Act; and
3. As a fallback, making KI available through the Strategic National Stockpile under Sections 127(a) and 127(d) of the Bioterrorism Act.

The discontinuation of NRC funding is conditioned on at least one of these options being in place before the November 2027 procurement deadline. If none are in place, the Commission has directed the staff to return with further options prior to the deadline.

Conclusion

The NRC's decision to end funding for KI stockpiles prioritizes science-backed protective actions and efficient resource allocation. The policy aims to maintain federal support for KI access as required by law, while emphasizing state and local responsibility for emergency planning

.

A former 23+-year member of the US Nuclear Regulatory Commission (NRC) staff whose tenure included numerous leadership roles, **Brooke Poole Clark** works with clients on a broad range of matters in the nuclear

industry. Brooke counsels clients on an array of NRC licensing issues related to operating and advanced reactors, the nuclear fuel cycle, radioactive materials licensing, and litigation before the Atomic Safety and Licensing Boards.
With a background in the nuclear power industry, **Timothy P. Matthews** represents and counsels electric utilities, nuclear industry suppliers, and other licensees before the Nuclear Regulatory Commission (NRC), the Department of Labor (DOL), and other regulatory agencies, and in US federal courts. He counsels clients in investigations, discrimination allegations, and regulatory compliance matters. He also advises on matters related to new nuclear plant development, electric utility industry restructuring, and complex disputes including mediation, arbitration, and litigation in US state and federal courts.

The political implications of Hiroshima's scarred "maidens"

By Alex Wellerstein
Source: https://thebulletin.org/2026/08/the-political-implications-of-hiroshimas-scarred-maidens/

Thirteen Hiroshima Maidens, scarred by the atomic bombing of Hiroshima, pause at a Hollywood church during their return to Japan after 18 months of plastic surgery in New York. Photo from the Los Angeles Public Library.

Aug 06 – On May 7, 1955, 25 young Japanese women arrived by airplane at Travis Airfield in California. They had flown there from Iwakuni Airbase in Tokyo a few days prior, with a layover in Honolulu. In California, a crowd greeted their arrival by showering them with flowers—and four women who, to the discomfort of many involved, used the opportunity to give a speech about the dangers of the hydrogen bomb.[1]
All of the women had been schoolchildren in Hiroshima on the morning of August 6, 1945, when the United States detonated an atomic bomb over the city. Many of them had been at work building firebreaks at the time of the bombing, and had all been severely burned by the attack. They were known in the press as the "Hiroshima Maidens," and they were part of a trip, organized by a Japanese reverend and an American journalist, and with the hesitant blessing of the US State Department, to bring the women to the United States, so their scars might be healed through the power of American plastic surgery. The primary intent was to ease their individual suffering, but the trip was understood by all involved to be more than that: the wounds being healed were not just physical, but political.

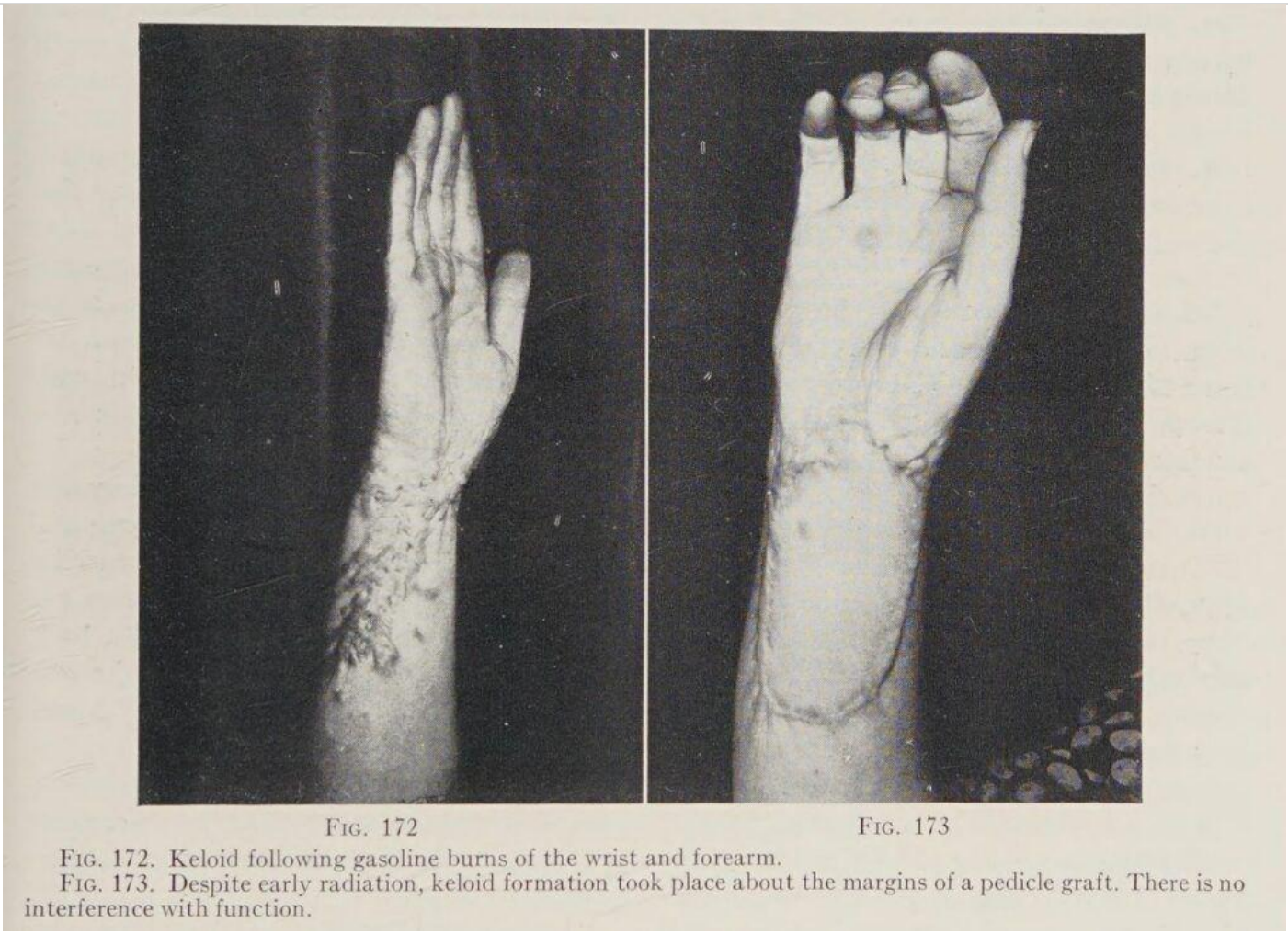


FIG. 172 FIG. 173

FIG. 172. Keloid following gasoline burns of the wrist and forearm.
FIG. 173. Despite early radiation, keloid formation took place about the margins of a pedicle graft. There is no interference with function.

Excerpt from Practice of Plastic Surgery, a foundational 1950 medical textbook written by Arthur Joseph Barsky.

Keloid scars are raised, thick, and irregular slabs of collagen. They can form after any injury, and can appear long after the injury itself. They generally do not hurt, but they can itch. They can also constrain muscle movement, decreasing mobility dramatically, depending on where they have formed. And they are disfiguring. For reasons that are apparently still not well-understood, such scars are more prone to develop in people of Asian, African, and Hispanic descent than in other peoples, and their heighted incidence among survivors of the atomic bombs has suggested that radiation exposure may increase the chances of their forming.

Photographs of the scarred bodies of atomic bomb victims are a common, if still controversial, accompaniment to discussions of the atomic attacks today. But in the immediate years after the end of World War II, they were much rarer. While an October 1945 story in *LIFE* magazine did feature two photographs of atomic victims, nearly all other popular photographs of Hiroshima and Nagasaki in the 1940s featured only photographs of the ruined buildings, not ruined bodies.

Less popularly, the bodies of victims were visible in official, technical publications on the "medical effects" of the bombs, with photographs of the victims months or even years after the bombings. These photographs were meant to display, cleanly and clinically, the effects of atomic bombs. They were not intended to provoke a sense of guilt, or even empathy, except as relates to how their audience might be interested in "effects data" for planning for the possible use of atomic bombs against themselves in the future.

Information about the survivors was tightly controlled within Japan by the occupation authorities, who feared that widespread discussion of the atomic bomb could lead to anti-American and pro-Communist sentiment.[2]

John Hersey's *Hiroshima*, published as a full-issue of *The New Yorker* in August 1946, was the first publication to really give international audiences a view of what the atomic bombings were like from the perspective of the bombed. One of the six survivors that Hersey profiled was an American-educated Methodist reverend, Kiyoshi Tanimoto. Tanimoto's account is gripping, and prominently features his inability to render the help to others that he wished that he could:

After crossing Koi Bridge and Kannon Bridge, having run the whole way, Mr. Tanimoto saw, as he approached the center, that all the houses had been crushed and many were afire. Here the trees were bare and their trunks were charred. He tried at several points to penetrate the ruins, but the flames always stopped him. Under many

houses, people screamed for help, but no one helped; in general, survivors that day assisted only their relatives or immediate neighbors, for they could not comprehend or tolerate a wider circle of misery. The wounded limped past the screams, and Mr. Tanimoto ran past them.

Hersey's account showed Tanimoto trying to help whom he could, but the magnitude of the disaster was overwhelming. Tanimoto himself suffered from a bout of radiation sickness, but recovered.

At the time of Hersey's book, there was no agreed-upon term to use for those people who were at Hiroshima and Nagasaki at the time of the bombings but who had not died. The term "survivor" was common in English usage, but the Japanese have tended to shy away from their equivalent (*seizonshai*), apparently finding it to contain a bit of guilt or blame in its focus on those who did not die. Conversely, the term "victim" appears to have been used more commonly in Japanese (*higaisha*) than in English, perhaps again because of its apparent connotation that the violence done to the Japanese was unjust. It is not clear exactly when the term *hibakusha*—"explosion-affected person"—became the preferred term; it seems like it was used in Japan by the mid-to-late 1950s, and became well-known in English only in the early 1960s.[3]

Kiyoshi Tanimoto stands between women who survived the Hiroshima bombing in May 1955. Photo from Japanese American National Museum.

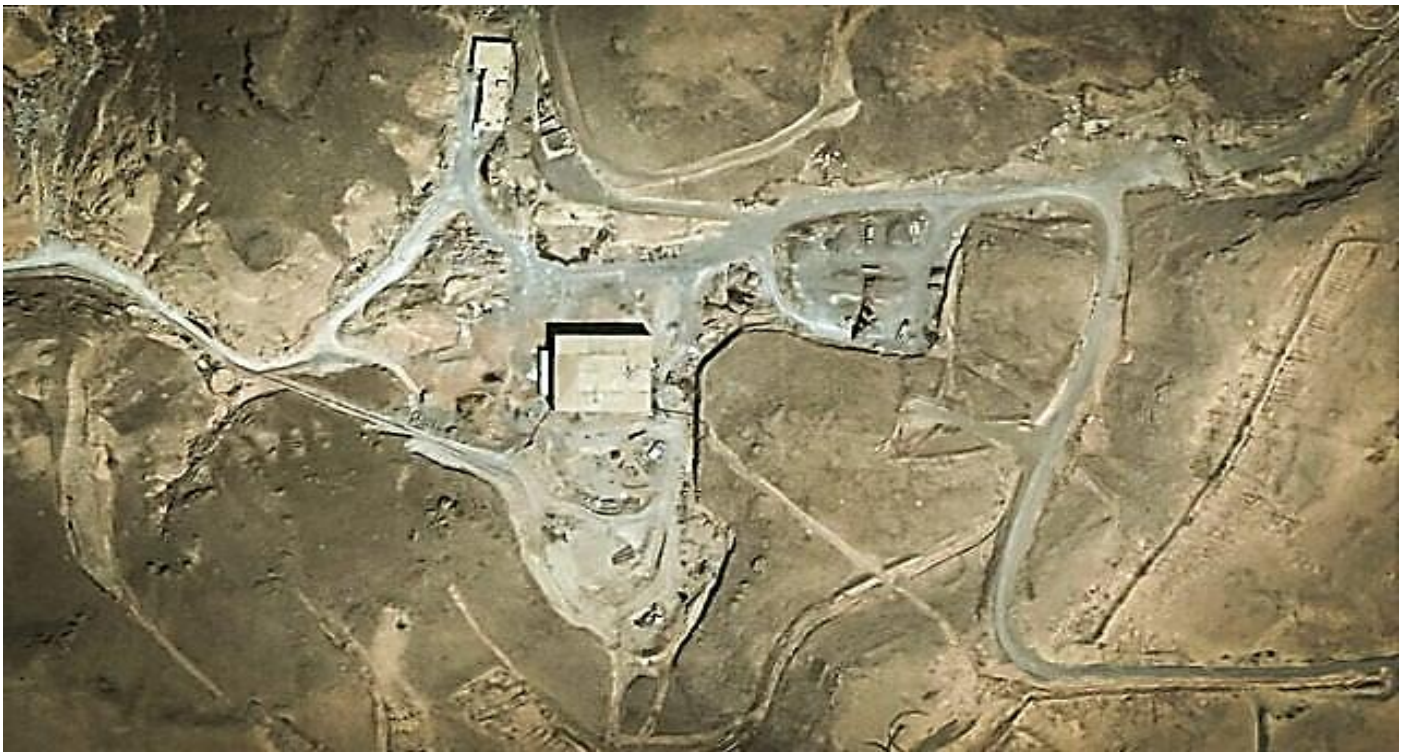
In the months after the bombing, his church destroyed, Tanimoto worked as a street preacher in Hiroshima. His message was both theological and political: Looking for someone to blame for the atomic bombing—whether the Americans who had dropped the bomb, or the Japanese military leaders who had started the war—was futile. Seek salvation through God alone. With parishioners, he began the work of rebuilding the church building, but the work was slow, for there was no money and few resources available. In 1948, he traveled to the United States to meet with an old classmate of his, in an effort to raise funds for a new church. While on the ship to San Francisco, he had an idea: He could dedicate his life to peace, and the rebuilding of his ruined Hiroshima would be the focal point for that. Without consulting anyone back home, he began to give sermons around the United States, taking up a collection for Hiroshima's rebirth. By the end of 1949, he had visited over 250 American cities, and raised about $10,000 for the church.[4] Tanimoto ultimately made a connection with the editor of the *Saturday Review*, Norman Cousins, who enthusiastically supported his plea, and ran an editorial proposing the creation of a World

Peace Center in Hiroshima. Although the article claimed to be in the name of the inhabitants of Hiroshima, no one but Tanimoto was ever consulted. A Hiroshima Peace Center Foundation was set up in New York by Cousins, in order to receive American donations.[5]
A few years later, as part of his peace center efforts, Tanimoto began a Bible study class for a group of women that he apparently called his "Keloid girls." He had offered other classes to victims of the bombing; the Keloid girls were special because their deformity and their sex was seen to demand a level of social exclusion that went beyond that of other Hiroshima victims. He would eventually petition the Hiroshima government for funds to give them plastic surgery, but was rejected. He then went to the Atomic Bomb Casualty Commission, the American-run effort to study the medical effects of the atomic bombings, but he was told that this was outside the commission's scope. The commission was not in the business of "treatment," and keloids were not unique radiation effects.[6]
Eventually, Tanimoto received some funding for surgery on the Keloid Girls within Japan, bringing a group of them to Tokyo in the spring of 1952, just after the end of the American occupation. There was also one male victim in the group, Kiyoshi Kikkawa. The Japanese press, to the surprise of Kikkawa, became completely focused on the women, dubbing them the *Genbaku Otome*: "Atomic Bomb Maidens." The women were attractive because of their symbolism of peaceful Japanese resilience. Kikkawa quickly felt it was going too far: At one point in their trip, the "maidens" were taken to tour a prison that held Japanese war criminals, and two of the "maidens" chosen as delegates proclaimed that while they had once held grudges against the Japanese militarists, they now saw them as being in the same situation as the survivors.[7] The surgeries were at least somewhat successful. Tanimoto had been in touch with Cousins about the "maidens" ideas, and with this pilot project completed—and its publicity value proven in at least the Japanese context—Cousins was ready to push for an American side of it. This would be larger and grander: more "maidens" (no men, this time), and the surgeries would be done in the United States, where plastic surgery was state-of-the-art. Cousins is credited for transforming their name: Having changed from "Keloid Girls" to "Atomic Bomb Maidens," they were now the "Hiroshima Maidens."[8]
For six months, Cousins sought funding for the project from foundations. It would be expensive: the transportation and lodging alone would cost tens of thousands of dollars, and the operations would drive the cost into the hundreds of thousands. He saw little success, apparently because the foundations blanched at the possible negative publicity that might attach to the project.[9]
Changing tactics, Cousins then attempted to see if the operations might be performed *pro bono*. Through his personal physician, he made a connection to Dr. Arthur Barsky, who ran and had founded the plastic surgery departments at Mount Sinai Hospital and Beth Israel Hospital in New York. Barsky had literally written one of the primary reference books for the field, *Principles and Practice of Plastic Surgery*, in 1950. Barsky had dedicated the book to "the soldier patients of World War II," on whom Barsky had practiced as a lieutenant colonel. The art of plastic surgery had long been associated with war injuries, and the "Maidens" project would be no exception.[10] Barsky agreed to the work, so long as he could be involved in the selection of proper candidates, as keloids were tricky, as his textbook explained. "The possibility of post-operative keloid formation is an ever present one," and certain parts of the body, and certain races of people, were more prone to them than others. "The cause and treatment of keloids are vexatious problems which have not been solved," he lamented. But experience had given promise that many could be treated, particularly with high doses of radiation, either directly to existing keloids or after the tissue had been excised.[11]
Barsky, along with his colleague William Hitzig — Cousin's physician — would fly to Japan to select the candidate "maidens," alongside Japanese physicians. The initial plan was to look only at women who were part of Tanimoto's Bible classes, but it was decided it would be more appropriate to widen the pool to any young woman who had been disfigured by the Hiroshima bombing. Although that would be a very large potential number, fewer than 50 actually showed up for the screening, perhaps because of rumors that the women would be subject to experimentation. Ultimately, 25 were deemed suitable candidates. The Mount Sinai surgeons had agreed to donate their labor, and the hospital agreed to provide bedding for little to no charge.[12]
After hearing about the project, Gen. John E. Hull, head of the US Far East Command, gave orders that a plane would be provided to bring the "maidens," Tanimoto, and the surgeons to the United States. The flight landed first in California, where it was met by journalists and a small protest, and then proceeded onward to New York. Before the operations would begin, there was publicity to attend to. Three days after landing New York, Tanimoto, and two of the "maidens" flew back to California to appear on the popular television show *This is Your Life*, with the goal of making a public appeal and raising perhaps a million dollars.

Alex Wellerstein is a Senior Fellow at the *Bulletin of the Atomic Scientists*, an associate professor at the Stevens Institute of Technology in Hoboken, NJ, and a visiting researcher at the Nuclear Knowledges program at the Center for International Studies at Sciences Po, Paris, France. His first book, *Restricted Data: The History of Nuclear Secrecy in the United States* was published in 2021 by the University of Chicago Press, and his second book, *The Most Awful Responsibility: Truman and the Secret Struggle for Control of the Atomic Age* was published by HarperCollins in late 2025. He is the author of the blogs Restricted Data and Doomsday Machines, and the creator of the NUKEMAP, an online nuclear effects simulation tool. He received a BA in History from the University of California, Berkeley, in 2002, and a PhD from the Department of the History of Science at Harvard University in 2010.

IAEA said prepping to remove Syria nuclear material left over from Assad era

Source: https://www.timesofisrael.com/iaea-said-prepping-to-remove-syria-nuclear-material-left-over-from-assad-era/



The Syrian nuclear reactor site before it was destroyed by the IDF in 2007, as seen by satellite (Photo: Google Earth)

Aug 11 – The International Atomic Energy Agency will soon remove nuclear material from clandestine facilities in Syria, after the White House reached understandings with Damascus and Israel, Axios reported on Tuesday.

The report, which cited Israeli and US officials, said a "dramatic, behind-the-scenes exchange of threats and diplomacy" erupted months ago concerning the secret site, and was only resolved several weeks ago with an agreement to remove the material, possibly by the end of this year.

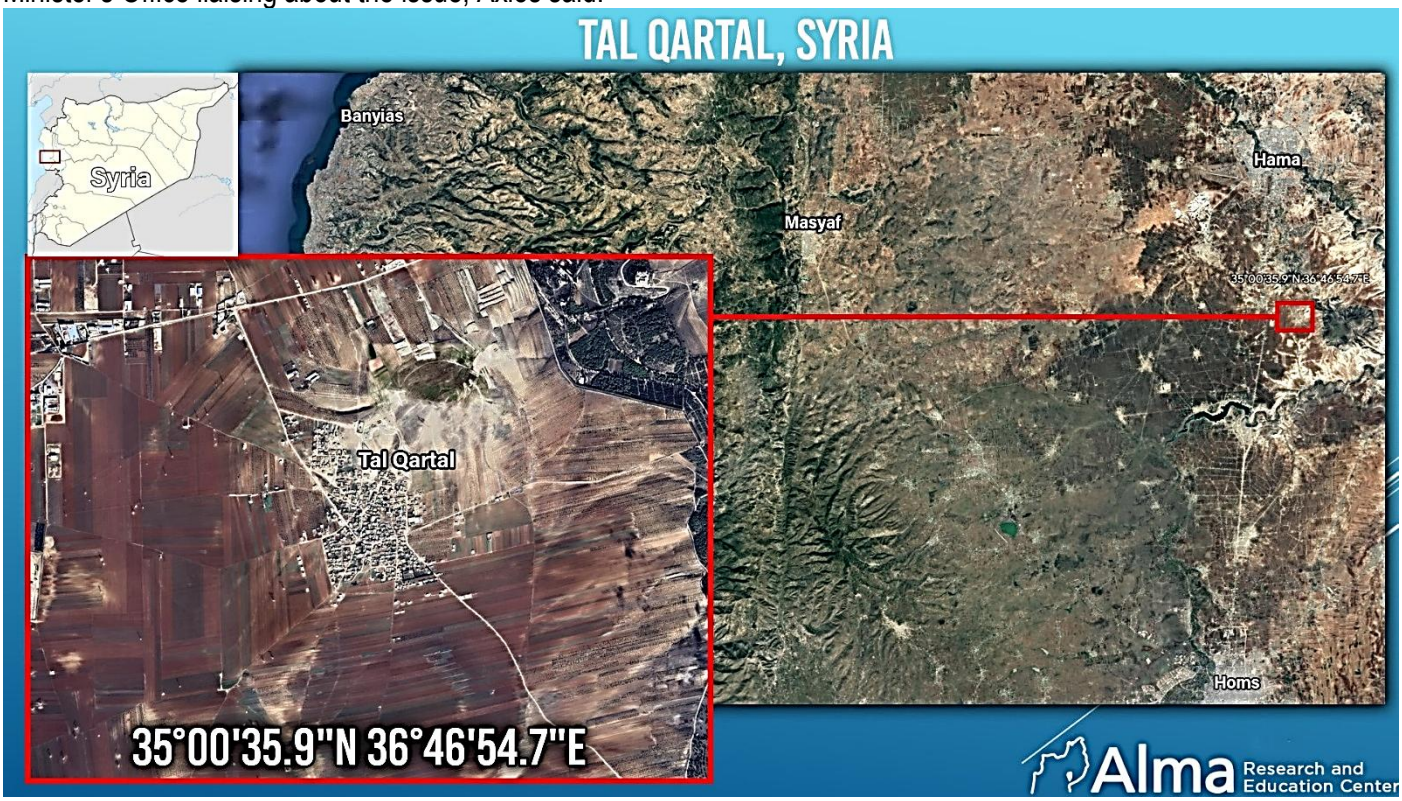
In 2007, Israel bombed the al-Kibar reactor in Deir Ezzor, Syria, where then-president Bashar al-Assad was — according to Israeli, US, and IAEA assessments — building a secret nuclear weapons program.

The attack, which Israel refused to comment on until 2018, is believed to have ended the regime's efforts to build a nuclear weapon. However, some research-and-development and storage sites remained intact, according to Axios.

Two Israeli officials told the outlet that a facility referred to as "Site 99" was used by the Assad regime to store nuclear material that remained in its possession after the al-Kibar reactor was destroyed.

The material included yellowcake, a uranium ore powder that can be used for a "dirty bomb" — a conventional bomb that also disperses radioactive material — or further enriched for a proper nuclear weapon.

At the end of 2024, when the Assad regime fell to the Islamist-led rebels who now head the Syrian government, Israel bombed the entrance to the al-Kibar site.
For more than a year afterward, Israel and the US discussed **Site 99,** with senior officials from the White House and Prime Minister's Office liaising about the issue, Axios said.



In June 2025, IAEA chairman Rafael Grossi said Syria had agreed to give inspectors access to suspected former nuclear sites. Israel threatened to strike the facility if the nuclear material were not taken out, or if the new government tried to reach it, Axios reported, though this threat was never close to being put into action, according to a US official.
According to the report, Israel and the US opted to work with the new Syrian government on the issue.
Several weeks ago, Israel detected activity at the site that led it to suspect the Syrian government was trying to access the material. Israeli officials worried that Turkey — a close ally of the new Syrian government — was helping Damascus to do so. According to Axios, the Trump administration then brought in the IAEA, which reached an agreement with the Syrian government that was signed three weeks ago.
A US official told Axios that Washington hopes to "kick off the removal process soon and end it before the end of the year."
The White House, IAEA, and PMO declined to comment on the report.

How the bomb failed to deliver security in Pakistan

By Sadia Tasleem
Source: https://thebulletin.org/2026/08/how-the-bomb-failed-to-deliver-security-in-pakistan/

Aug 11 – The world is witnessing a renewed interest in nuclear weapons. From nuclear modernization plans in the United States, to growing anxieties in Europe and East Asia about the reliability of the US security umbrella, to discussions in South Korea, Canada, and elsewhere about acquiring nuclear weapons, the thinking about nuclear weapons appears to be shifting. But Pakistan shows something different and unexpected. In the aftermath of a short but limited three-day war with India in May 2025, Pakistan's strategic elite had to face a hard truth: Nuclear weapons have limits. They do not prevent every crisis or escalation of military conflict, and they do not deter coercion. This is a reality that nuclear weapon states like the United States, Russia, and the United Kingdom have also experienced but rarely acknowledged publicly. Evidence for this realization comes from how Pakistan commemorated the first anniversary of the military confrontation this May (the *Marka-e-Haq*) and the anniversary of the day in

May 1998 when Pakistan conducted its first nuclear tests (the *Youm-e-Takbeer*). Those commemorations played down the importance of nuclear weapons, in part because the May 2025 conflict highlighted the limitations of Pakistan's nuclear arsenal and the risks it can impose during a crisis.

Little fanfare

Marka-e-Haq has been elevated into a visible state narrative since May 2025. The first major indication of this came in July 2025, when the government announced that Pakistan's Independence Day would be celebrated under the theme of Marka-e-Haq, encouraging participation from state institutions and organizations across the country, including educational institutions. By the first anniversary in May 2026, the commemoration had become even more visible. Coordinated celebrations were organized across government departments, provincial governments, cultural institutions, public universities, diplomatic missions, and the armed forces. It clearly reflected the importance attached to the event by the state. By comparison, Youm-e-Takbeer passed with remarkably little fanfare. There were no large-scale celebrations, and no sustained media attention. Senior civilian and military officials issued brief statements on social media, but the day lacked the symbolic prominence that had accompanied it during previous periods of center-right Pakistan Muslim League rule. This was particularly striking given that the party, which currently heads the civilian government, has long taken pride in its decision to conduct Pakistan's nuclear tests in May 1998.

Even more telling were the visual cues.

Islamabad witnessed massive public displays of conventional military achievement, portraits of military leaders and narratives of battlefield success. From public transport emblazoned with images and slogans of victory, to posters lining major roads, to the naming of a public square, the message was impossible to miss. Notably, however, nuclear weapons and nuclear-capable missiles were conspicuously absent from these images of militarization. This was not because the state lacked the opportunity to showcase them. It was a deliberate effort to communicate a different message.

When the bomb loses relevance

The reason has to do with the dominance of the military in all aspects of Pakistan. The political interests of the military and its institutions are better served by conventional military success than nuclear weapons. The bomb's destructive power appears largely self-contained; its existence does not easily translate into visible demonstrations of military competence or leadership. Moreover, Pakistan's nuclear program was not solely the achievement of military institutions. The program was populated largely by scientists and engineers, and the decision to test nuclear weapons in 1998 was ultimately taken by a civilian government.

Even though the bomb has served the military's political objectives in many ways in the roughly three decades since those tests, by reinforcing its claim to be the indispensable guardian of national security, legitimizing its dominant role in strategic decision-making, and enhancing its institutional authority at home and abroad, its ability to stoke pro-military sentiment among the masses remains limited. For a military establishment emerging from one of the most serious legitimacy crises in Pakistan's history, highlighting conventional military achievements therefore has much more political value than celebrating the bomb. The former reinforces the image of the armed forces as the nation's active guardian. The latter does not. This shift also reflects a deeper problem. For a long time, Pakistan's strategic elite has projected nuclear deterrence as a nearly absolute guarantee of national security and territorial sovereignty. The promise, which goes by the name full-spectrum deterrence, was huge but false: Nuclear weapons would prevent aggression at all levels of conflict spectrum. But recent events have exposed the limitations of those claims. Nuclear deterrence did not prevent the May 2025 crisis. The fear of escalation did not keep India from striking deep in Pakistan's heartland. Nor did it provide a decisive solution to the challenges Pakistan confronted. Unlike previous episodes, these limitations have become increasingly difficult to obscure. The gap between the promises associated with nuclear deterrence and the realities of strategic competition has become too visible to ignore.

What we see today is the replacement of nuclear weapons at the apex of military strategy with a greater emphasis on conventional militarism; this is precisely the reason why Pakistan sees a spike in its defense spending coupled with the establishment of a new Army Rocket Force Command. The visibility of the senior leadership of the Rocket Force in international defense forums such as the Shangri-La Dialogue can be interpreted as signaling Pakistan's conventional military preparedness. Nuclear weapons seem to have been displaced from the center stage.

Limited utility, profound risks

The May 2025 crisis exposed another serious challenge. Nuclear weapons were not merely absent as a solution; instead, they emerged as part of the problem. Nuclear weapons did not deter the crisis from unfolding, nor did they offer policymakers a way out once hostilities escalated. Instead, the presence of nuclear weapons constrained Pakistan's response options by introducing constant risk of misperception and inadvertent escalation. For example, Pakistan's response to India's missile strikes had to account not only for military effectiveness but also for how its actions might be interpreted by New Delhi. Any missile launch carried the possibility that India could mistake a conventionally armed strike for the opening phase of a nuclear attack, increasing the dangers of rapid

escalation. The challenge was not simply how to respond, but how to respond without triggering catastrophic miscalculation. The risks were further exacerbated by the reports during the final stages of the confrontation that Pakistan had convened—or was preparing to convene—its National Command Authority, shortly before the United States announced a ceasefire. Whether the meeting actually happened or not is less important than what its likelihood revealed: The crisis had reached a point when nuclear command and control entered the public discourse, narrowing the margin for error and increasing the urgency of de-escalation. International diplomatic intervention helped bring about a ceasefire before these risks increased. Had the fighting continued, the possibility of an unintended or uncontrolled escalation could not have been dismissed.

If anything, this experience reveals that the utility of nuclear weapons is narrower than many of their advocates claim. Pakistan's experience is worth studying not because it explicitly disproves deterrence, but because it demonstrates the dangers of treating nuclear weapons as a universal solution to security problems.

Countries debating whether to acquire nuclear weapons often ask what the bomb can do for them. Pakistan's recent experience suggests they should also ask what military constraints and risks the bomb introduces.

Sadia Tasleem is a lecturer with the Department of Defense and Strategic Studies at Quaid-i-Azam University, Pakistan.

PERSPECTIVE Open Access
Military Medical Research (2018) 5:27

Medical management of victims contaminated with radionuclides after a "dirty bomb" attack

Alexis Rump*, Benjamin Becker, Stefan Eder, Andreas Lamkowski, Michael Abend and Matthias Port

Abstract

A wide spectrum of scenarios may lead to radiation incidents and the liberation of radioactive material. In the case of a terrorist attack by a "dirty bomb", there is a risk of mechanical and thermal trauma, external irradiation, superficial contamination and incorporation of radioactive material. The first treatment priority must be given to the care of trauma patients with life-threatening injuries, as the health effects of radiation occur with latency. Radionuclide incorporation will lead to a longer-lasting irradiation from inside the body, associated with a higher risk of stochastic radiation effects (e.g., occurrence of tumors) in the long run. It must be expected that victims with potentially incorporated radionuclides will far outnumber trauma patients. The elimination of radionuclides can be enhanced by the administration of decorporation agents such as (Ca) Diethylenetriaminepentaacetic acid (DTPA) or Prussian blue, reducing the radiological burden of the body. There is still no consensus whether decorporation treatment should be started immediately based only on a suspicion of radionuclide incorporation ("urgent approach") or if the results of internal dosimetry confirming the necessity of a treatment should be awaited, accepting the delay caused by the measurements and computations ("precautionary approach"). As the therapeutic effectiveness may be substantially decreased if treatment initiation is delayed only by several days, depending on the radionuclide, the physicochemical properties of the compounds involved and the route of absorption, we favor an "urgent approach" from a medical point of view. In doubt, it seems justified to treat victims by precaution, as the adverse effects of the medication seem minimal. However, in the case of a high number of victims, an "urgent treatment approach" may require a large number of daily doses of antidotes, and therefore, adequate investments in preparedness and antidote stockpiling are necessary.

Keywords: Medical NRBC protection, Radiological emergency, Dirty bomb, Combined injuries, Radionuclide incorporation, Decorporation therapy

Analyst weighs Guam's risks as Pentagon reviews nuclear options

By Walter Ulloa | The Guam Daily Post

Source: https://www.postguam.com/news/local/analyst-weighs-guam-s-risks-as-pentagon-reviews-nuclear-options/article_6cece16f-9ee7-418f-9139-0543a70effa7.html

Aug 17 – A Chinese nuclear strike on Andersen Air Force Base would likely not stay confined to the base. Radioactive fallout would probably drift across Guam's civilian neighborhoods, too, according to a local security analyst who is watching a Pentagon review that could put smaller nuclear weapons back into everyday U.S. war planning.

Leland Bettis, director of the Pacific Center for Island Security, said the island's geography works against it in that scenario.

"I'm not an expert in fallout, but the prevailing winds in Guam do come from the northeast," Bettis said in an interview with the Post. "So if Andersen would hit, there'd be a lot of fallout spread across the island."

Bettis based that scenario on China's DF-26, a missile capable of carrying either a conventional or nuclear warhead. He said a conventional strike from its roughly 3,000-pound payload alone would be enough to destroy a launch site, an airfield or a port facility. If the warhead were nuclear instead, he said, it might have a relatively small warhead aimed at Andersen or the port.

That is not a scenario Bettis raised on his own. It tracks with a shift Pentagon officials confirmed earlier this month.

Undersecretary of Defense for Policy Elbridge Colby confirmed the Pentagon is reexamining how and when the United States might use nuclear weapons, telling an audience in Omaha, Nebraska, on Aug. 5 that the goal is to give the White House more choices in a crisis, though he said no changes have been publicly finalized.

"There's, you know, a review or sort of relook at the employment guidance, which is more about the commander of (Strategic Command) and his responsibilities and the secretary's authorities, etcetera," Colby said, according to Breaking Defense, which obtained audio of his remarks.

Adm. Richard Correll, head of U.S. Strategic Command, told reporters afterward that the goal is to widen the range of choices available to the president in any escalation scenario, from conventional weapons to a limited nuclear response, without changing the country's declared nuclear policy.

According to Breaking Defense, the push builds on weapons already added to the arsenal for that purpose. The Trump administration revived a submarine-launched cruise missile known as the SLCM-N during its first term, and the Biden administration later approved a higher-yield version of the B61 gravity bomb, the B61-13, giving commanders options below the size of the country's largest warheads.

Bettis said the review fits a broader shift already underway in how the Pentagon views Guam. For decades, he said, the island felt insulated because Cold War strategy rested on the idea that using a nuclear weapon would doom the attacker as much as the target. He said that calculus has changed.

"Under the old environment, we had mutually assured destruction," Bettis said. "We had assurances that basically we were going to be safe as long as everyone else was safe. Well, that's really no longer the case."

He said the dynamic now amounts to islanders shouldering a burden while the outcome benefits everyone around them, without local leaders demanding anything in return.

"Right now, all we really are is the deductible on a U.S. insurance policy to keep conflict limited to the region," Bettis said. "We are a cost in that policy. We're the deductible."

Bettis also pointed to a quieter signal: the military's retreat from an earlier vision of homeporting a full squadron of next-generation submarines on the island.

"If you go back two or three years, that is no longer the case," Bettis said. He said Guam today hosts four Los Angeles-class submarines and one Virginia-class submarine, while the Navy has shifted its buildout toward Western Australia as part of the AUKUS agreement with Australia and Britain. "All the plans for Polaris Point have been significantly downgraded," he said.

Bettis has also been tracking Chinese research vessels sailing close to Guam, activity he says fits a broader effort to locate the acoustic signatures of American undersea forces.

"It's doing a lawnmower pattern, which is typical of oceanographic survey work," Bettis told ABC Australia in reporting published last week, describing the movements of the Chinese research vessel Ji Di. "The fact it's doing the lawn

mowing positioning directly off of Guam where U.S. submarines are based underscores the potential for some of the dual use capabilities."

Jennifer Parker, a former Australian anti-submarine warfare officer now at the University of Western Australia, said understanding the ocean floor helps navies both hide and locate vessels because acoustic signals move differently depending on depth and temperature.

"The primary way you detect subs is through sound, passive by listening or active by putting sound in the water and getting it to bounce back," Parker said.

The pattern is not confined to Guam's waters. The same vessel, an icebreaker called the Ji Di, spent 14 hours moving through Palau's exclusive economic zone in late May, coming within 74 kilometers of the coastline. Palau's president, Surangel Whipps Jr., said Chinese ships have entered his country's waters without permission and given no explanation for their presence.

"Their research vessels come into our EEZ uninvited and do whatever they want," Whipps said. "What are they looking for, and what is the ultimate intention?" Pacific Islands Forum foreign ministers, meeting last week, agreed to push for stronger nuclear language in this year's leaders' declaration and asked officials to draft a joint statement on the missile test for consideration at next month's forum in Palau. The move followed China's ballistic missile launch into the Pacific in July, its first such test since September 2024.

Guam's senators have pushed for more openness from military commanders on related questions. Sen. Telo Taitague, who chairs the legislative committee overseeing the buildup, plans another informational briefing next month with Del. James Moylan, Joint Region Marianas and Joint Task Force Micronesia. U.S. Indo-Pacific Command has so far declined to say whether the Army's Dark Eagle hypersonic missile system remained in the Marianas following last month's Valiant Shield exercise, citing operational security.

Bettis said the more pressing task for residents is not any single weapons system but understanding how the pieces - tactical nuclear weapons, missile stockpiles, submarine posture and Chinese surveillance activity - connect to one another.

"We need to understand it," Bettis said. "That's the bottom line. We need to understand it if we're going to be able to deal with it."

Pakistan's Nuclear Arsenal is Becoming More Vulnerable

By Sara Harmouch

Source: https://www.irregularwarfare.org/pakistans-nuclear-arsenal-is-becoming-more-vulnerable/

Aug 21 – "We call on the Pakistani army and security services not to obey Pakistan's corrupt political and military leadership and urge them to overthrow it," al-Qaeda declared in an April appeal to Pakistan's military amid the Afghanistan-Pakistan war. The same statement told Pakistani pilots to defy their commanders and target them "by any means possible." This appeal was aimed at members of the military and security services entrusted with protecting one of the world's fastest-growing nuclear arsenals.

For more than twenty years, several barriers kept Pakistan's nuclear weapons beyond militant groups' reach. The United States and its allies had troops and intelligence networks in Afghanistan. Washington and Islamabad cooperated on counterterrorism and nuclear security. Pakistan also built a serious system to guard its arsenal. Together, these measures made it much harder for al-Qaeda to gain access to Pakistan's nuclear enterprise.

Those barriers are now weaker, and the Afghanistan-Pakistan war is becoming a nuclear security problem. Al-Qaeda's enduring nuclear ambition now coincides with greater access through regional militant networks, less U.S. visibility, and growing pressure on the very institutions Pakistan relies on to secure its arsenal. Afghanistan is again a sanctuary for al-Qaeda, this time under a Taliban government rather than an insurgency operating on the margins. The

United States no longer watches the threat from inside Afghanistan. Al-Qaeda is also strengthening its ties with the Pakistani Taliban, Tehrik-e Taliban Pakistan (TTP), which is already at war with Islamabad. Pakistan, meanwhile, is guarding a larger and more mobile arsenal while fighting terrorism, insurgency, and a cross-border conflict along its western frontier. The war now reaches beyond the border and into the security of Pakistan's nuclear arsenal.

Danger Contained

Al-Qaeda's interest in weapons of mass destruction is not new. In the 1990s, Osama bin Laden called acquiring nuclear weapons a religious duty. His operatives shopped for uranium and forged ties with Pakistani nuclear scientists. After 9/11, American forces found nuclear documents in al-Qaeda safe houses and uncovered meetings between the group's leaders and senior figures from Pakistan's nuclear program, including Bashiruddin Mahmood and Chaudhry Abdul Majeed. Both men were tied to Ummah Tameer-e-Nau, an organization founded by retired Pakistani nuclear scientists and military officers. These efforts show that al-Qaeda has long looked to scientists, military personnel, and procurement networks for access to nuclear knowledge and material. Pakistan's own A.Q. Khan network demonstrated that designs, technology, and components could move through Pakistani channels despite formal controls.

By 2007 and 2008, the threat had moved closer to Pakistan's nuclear geography. The Pakistani Taliban (TTP) and al-Qaeda struck installations tied to, or near, Pakistan's nuclear complex. The most significant attack came in 2008, when the Pakistani Taliban attacked the Wah cantonment, described as one of Pakistan's main nuclear-weapons assembly sites. That same year, a bipartisan U.S. commission warned that if you mapped terrorism and weapons of mass destruction, "all roads would intersect in Pakistan." A 2013 House Homeland Security hearing reached a similar conclusion: al-Qaeda had pursued nuclear weapons for two decades, and terrorists had repeatedly attacked facilities associated with Pakistan's nuclear enterprise.

Yet the breach many feared never occurred because sustained U.S. and Pakistani pressure made it difficult for al-Qaeda to turn its nuclear ambitions into access. American and NATO forces were still in Afghanistan, while U.S. intelligence maintained access to the borderlands, sources inside jihadist networks, and the ability to disrupt plots early. Washington and Islamabad cooperated on counterterrorism, and the United States spent roughly $100 million helping Pakistan secure its nuclear arsenal. Michael Morell, the CIA's former acting director, later said American intelligence may have been what prevented al-Qaeda from acquiring a Pakistani nuclear weapon.

Pakistan, for its part, was not careless. After the Khan scandal and post-9/11 insider revelations, it built a serious custodial architecture: the Strategic Plans Division to manage the arsenal, a dedicated security force, a personnel reliability program to screen custodians, and a posture that kept warheads disassembled and stored apart from their delivery systems. Violence from 2007 to 2009 tested that system, but the barriers held. They held because pressure, visibility, U.S.-Pakistan cooperation, and Pakistani institutional effort reinforced one another. Today, that system is operating under greater pressure and with less outside support.

The Buffers are Weakening

Al-Qaeda's nuclear ambitions never died. In 2009, an al-Qaeda leader in Afghanistan said the group would use Pakistan's nuclear weapons against the United States if it could. More than a decade later, the theme resurfaced. In 2023, Ibrahim al-Qosi, a senior al-Qaeda in the Arabian Peninsula figure, released a revised edition of his 2021 book with a new chapter on weapons of mass destruction, signaling renewed emphasis inside al-Qaeda's Yemen branch. That same year, Saif al-Adel, al-Qaeda's presumed de facto chief, wrote that only weapons of mass destruction could deter the West. Al-Adel helped run al-Qaeda's original nuclear effort and its early attempts to buy uranium. The group also retains some of its 9/11 era nuclear experts, including Ali Sayyid Muhammad Mustafa al-Bakri, the chemist Khalid Sheikh Mohammed called al-Qaeda's "nuclear chief," who remains at large in Iran alongside al-Adel. The ambition has endured, but the environment around it has changed.

The first buffer to weaken was the pressure al-Qaeda faced in Afghanistan. After the U.S. withdrawal and the Taliban's return to power, Washington had to watch from over the horizon, with less visibility into the networks it once kept under pressure. Al-Qaeda also regained a permissive sanctuary inside a Taliban-run state—something it had not enjoyed for two decades. From there, the group is rebuilding training pipelines, deepening regional relationships, and tightening operational ties to the TTP, the Pakistani Taliban insurgency fighting Islamabad. It is increasingly acting as a service provider to the TTP, offering guidance, expertise, and operational support. That gives al-Qaeda an indirect pathway into Pakistan through a movement already at war with the state and already attacking security institutions. In that way, al-Qaeda brings the nuclear ambition while the TTP brings local access, manpower, and direct pressure on the institutions Pakistan relies on to secure its arsenal.

The second buffer to weaken was the U.S.-Pakistan partnership that once helped contain that danger. Washington suspended broad security assistance to Pakistan in 2018, and cooperation with Islamabad is now narrower, colder, and more transactional. The contrast with the post-9/11 period is sharp. In 2022, President Biden called Pakistan "one of the most dangerous nations in the world," citing "nuclear weapons without cohesion." By 2026, Washington no longer treated Pakistan's nuclear enterprise as a joint

security problem to manage with U.S. assistance. It treated it as a growing source of concern.

Pakistan, meanwhile, faces heavier strain than at any point in that period. It ranks among the countries most affected by terrorism. It is fighting the TTP in Khyber Pakhtunkhwa, Baloch separatists and ISIS-K in Balochistan, and now a cross-border war along the same western frontier. In March 2026, as the war escalated, Pakistan accused Afghan forces of launching a drone attack in Rawalpindi, the headquarters of the Pakistani army. The significance is not only symbolic. Much of Pakistan's nuclear weapons infrastructure sits in the north and west of the country and around the Islamabad-Rawalpindi corridor. The conflict now touches the military and security geography that underpins Pakistan's nuclear establishment.

Pressures on Pakistan's Nuclear System

These pressures are reaching Pakistan's nuclear command and security system at a difficult moment. A 2025 constitutional amendment elevated the army chief into the new role of Chief of Defense Forces, abolished the old Chairman Joint Chiefs of Staff Committee post, and created a new Commander of the National Strategic Command with nuclear-command responsibilities. Critics argue that this kind of redesign breeds uncertainty exactly where Pakistan needs clarity: who coordinates the services, who manages nuclear command, and how authority flows in a crisis. The army and intelligence services entrusted with the arsenal are the same institutions absorbing the country's internal and border pressures, with less external visibility than at any point since 9/11.

The arsenal itself has changed too. Pakistan now fields a larger and more mobile arsenal of roughly 170 warheads, with more likely if current trends hold. Its nuclear architecture has diversified, with nuclear-capable aircraft, cruise missiles, and road-mobile ballistic systems, including the Shaheen-II. The short-range Nasr, a road-mobile, low-yield nuclear delivery system built for quick response, captures the security problem mobility creates. More dispersed systems mean more weapons, launchers, personnel, and command arrangements to protect, especially in a crisis.

The danger today is not that al-Qaeda will steal a Pakistani nuclear weapon next week. Pakistan reportedly guards its warheads through a layered custodial system, separates them from delivery systems, and protects them with use-control measures. Nor is there evidence that any terrorist group has ever produced fissile material. The more plausible danger lies around the arsenal: a compromised officer, a sympathetic technician, an insider at a sensitive facility, a militant network with access to security personnel, or a system under enough strain to miss the warning signs.

Al-Qaeda has looked for openings like these before. Pakistani nuclear scientists met with bin Laden and Ayman al-Zawahiri in Afghanistan, where one of them briefed the al-Qaeda leaders on the infrastructure a nuclear weapons program would require. The A.Q. Khan network ran for years before its exposure, showing that insiders can move nuclear designs, components, and procurement channels without immediate detection. In 2014, al-Qaeda recruited Pakistani naval officers for an alleged plot to seize a frigate and turn its missiles on U.S. vessels in the Arabian Sea; the officers had boarded the warship before authorities stopped them. Each case shows al-Qaeda trying to penetrate institutions outsiders often treat as impermeable.

Nuclear security ultimately depends on people, and al-Qaeda is now openly courting them. Its April statement is a recruitment pitch aimed at Pakistani soldiers, pilots, intelligence officers, and security personnel, the people every custodial system relies on. By urging them to side with their "Afghan brothers" and restore Pakistan to what it called its "honorable past," the group appeals to loyalty, grievance, religious identity, and obedience inside the institutions that hold the state together and protect its arsenal. Those appeals land in institutions with known vulnerabilities. Overlapping currents of anti-U.S. sentiment, Islamist sympathy, personal and clan ties to militants, and resentment over campaigns against fellow Pakistanis have long run through parts of Pakistan's army and intelligence services. The concern is not hypothetical. Hamid Gul, a former director-general of Pakistan's Inter-Services Intelligence agency, was implicated for advising al-Qaeda and the Taliban on developing chemical, biological, and nuclear weapons. These are the vulnerabilities al-Qaeda is now trying to widen. They point to the scenario Bruce Riedel warned about: Pakistan's arsenal remains under military control, but if militant Islamists and al-Qaeda took control of the country, "the arsenal would fall with it."

Prevention is the Only Option

Al-Qaeda now operates closer to the people and networks that could give it access to Pakistan's nuclear enterprise. It has sanctuary under Taliban rule, deeper ties to Pakistani militants fighting Islamabad, and a message aimed directly at soldiers, pilots, intelligence officers, and security personnel. Pakistan is fighting multiple threats while guarding a larger and more dispersed arsenal. The United States, meanwhile, has less visibility into the networks that could turn intent into opportunity. Together, these pressures narrow the margin for error.

That is why the Afghanistan-Pakistan war cannot be read only as a border conflict or regional crisis. It is pressing on the institutions that secure Pakistan's nuclear arsenal. In May, shortly after al-Qaeda's appeal, the United Nations warned that the risk of nuclear terrorism was higher than ever. The 2026 U.S. Counterterrorism Strategy calls keeping unconventional weapons out of terrorists' hands a "no-fail" mission.

Pakistan is now one of the places where that mission will be tested.

Washington does not need to reoccupy Afghanistan, and it should not pretend the Taliban is a counterterrorism partner. But it can take less cinematic and more effective steps: restore intelligence visibility where possible, revive the nuclear security dialogue with Islamabad, and press regional capitals with leverage over Kabul and Islamabad to treat the war itself as a nuclear security concern. Above all, Washington should recognize that deterrence does not cover this gap. A weapon or a quantity of fissile material passed through an insider or a militant network with no territory and no capital may arrive without a return address. By the time responsibility is established, if ever, the damage would already be done. The only workable strategy is to prevent access before material, expertise, or custody slips outside state control.

For twenty years, the "loose Pakistani bomb" was feared yet never materialized. That record should not reassure Washington. It should remind Washington what prevention achieved when pressure, visibility, cooperation, and Pakistani security measures worked together. The Afghanistan-Pakistan war is now weakening those conditions, one barrier at a time. We have spent two decades not failing at this. The catastrophe would be to assume that is good enough.

Dr. Sara Harmouch is a *counterterrorism expert* and CEO of H9 Defense, a research and advisory consultancy specializing in terrorism, national security, geopolitical risk, threat assessment, and open-source intelligence. Her research examines militant organizations, alliance dynamics, irregular warfare, and the conditions under which armed groups adapt, endure, and expand from domestic to international violence. She also studies terrorism's intersection with emerging technologies, unconventional capabilities, and weapons of mass destruction. Dr. Harmouch is a Security Fellow at the Frontier Security Institute and a Non-Resident Fellow at George Washington University's Program on Extremism. Raised in Lebanon and a native speaker of Arabic and French, she conducts fieldwork and primary-source research across the Middle East, Africa, Asia, and Europe. She regularly briefs U.S. and international security agencies, and her work has appeared in The Washington Quarterly, Lawfare, War on the Rocks, Defense One, The Long War Journal, NPR, and VOA, among others. She holds a Ph.D. in Justice, Law, and Criminology from American University and a master's degree in International Relations, Science, and Technology from the Georgia Institute of Technology.

ICI International CBRNE INSTITUTE

CBRNE-Terrorism Newsletter

C²BRNE DIARY

Weapons, military equipment and

EXPLOSIVE NEWS

Russia is attacking Ukraine with radioactive munitions—new cases have been confirmed in the Chernihiv region

By Serhiy Malamura

Source: https://news.liga.net/en/war/news/russia-is-attacking-ukraine-with-radioactive-munitions-new-cases-have-been-confirmed-in-the-chernihiv-region

Examination of the debris (Photo: SBU)

July 23 – Russia is using depleted uranium munitions in its attacks on Ukraine. These hazardous components were found, in particular, in Geran-2 strike drones, which the enemy used to attack the Chernihiv region in April–June 2026. This information was reported to by, the Security Service of Ukraine. Ukraine and its international partners report hundreds of instances in which Russian troops have used grenades containing irritants and toxic substances, as well as other prohibited weapons. At the same time, Moscow accuses Ukraine of violating the international ban on weapons of mass destruction. To learn how far the Kremlin might go in its pursuit of a phantom "chemical threat"— —read the article at .

The Russians are using radioactive elements in their R-60M missiles, which they have adapted for use on their UAVs. The radiation levels in the examined combat units of Russian drones exceed the norm by 80 times, and in some cases even more.

"The gamma radiation levels of individual fragments of enemy drones ranged from 8.3 to 24 μSv/h, compared to the permissible health standard of 0.3 μSv/h. Such radiation levels pose a direct threat to human health and the environment," the SBU noted.

The expert analysis determined that the warheads of these missiles contain uranium-235 and uranium-238. Russia's use of radioactive components has been confirmed by dosimetric equipment and a mobile automated radiological monitoring system.

After being examined, the radioactive munitions were neutralized to ensure public safety. SBU investigators have opened a criminal case under Article 438 of the Criminal Code of Ukraine (war crimes).

- In April, Russia had already attacked the Chernihiv region with drones emitting elevated levels of radiation.
- In July, the SBU documented new evidence of the enemy's use of radioactive weapons in the Sumy region.

Mystery surrounds identity of man killed by his own pipe bomb in Dublin hostel

Source: https://www.irishtimes.com/crime-law/courts/2026/07/29/mystery-surrounds-identity-of-man-killed-by-his-own-pipe-bomb-in-dublin-hostel/

July 29 – A man killed when a pipe bomb he had made exploded in his room in a Dublin homeless hostel cannot be identified, two decades after his arrival in Ireland and more than two years since his death. The man, known as Igor Dmitrov, is believed to have been aged 34 when he died at the Depaul

hostel on Little Britain Street in Dublin's north inner city in January 2024. He arrived in Ireland in late 2006 on a passport now known to have been fake. He secured a personal public service number and identification documents, using the name and other details on the passport. He continued to live under that identity, accessing homeless services, social welfare payments and medical treatment. Det Garda Shane D'Arcy from Dublin's Bridewell Garda station told a coroner's inquest the passport's number and other details were from the passport of a Lithuanian national that was long reported lost. Those details were used to create a new, fraudulent passport for the dead man, who presented himself as Igor Dmitrov, a Lithuanian born on June 12th, 1985. D'Arcy told Dublin Coroner's Court that when it was established the passport had been created using another person's details, other efforts were made to establish the identity of the dead man. His phone was analysed, revealing he had been in touch with one man abroad, based in Germany. The man was contacted and said he knew the deceased as "Igor", but from Ukraine, not Lithuania. Efforts were also made to contact a woman said to be the sister of the deceased, but she could not be found. When Coroner Dr Clare Keane asked D'Arcy whether it was correct to say the dead man could not be identified, he replied: "That's correct." The inquest was told the dead man had substance abuse issues from the age of 13 and also suffered from schizophrenia.

A narrative verdict was recorded, stating the man "known as Igor Dmitrov" died of traumatic blunt force injuries following the explosion of a device, which was next to his head at the time.

Dr Joanne Fenton, a consultant psychiatrist, said while Dmitrov had a long history of mental illness, he did not appear to be suicidal during his interactions with medical staff, up to the period before his death.

However, Det Garda Seamus O'Donnell, a Garda ballistics expert, said the pipe bomb made by Dmirov was placed on his bed where a pillow would be expected to be found.

He said the back of the dead man's head was next to the bomb, or he was lying on top of the bomb, when it exploded, adding that he felt the man's positioning was "significant".

O'Donnell told the inquest the bolts used at the ends of the bomb had been bought online and that a fuse, similar to one found in fireworks, was used. While black powder or "old-fashioned gun powder" could be used for the explosives, cap gun caps were also found in the room, which was destroyed by the explosion and the fire it caused, the inquest heard.

Manhunt after car carrying explosives crashes in police chase

Source: https://www.dutchnews.nl/2026/07/manhunt-after-car-carrying-explosives-crashes-in-police-chase/

July 29 – Police are trying to trace the occupants of a car that

crashed at high speed in Amsterdam while carrying high explosives, causing dozens of homes to be evacuated. The black BMW M1 was being pursued by Dutch police after their German colleagues spotted it at the scene of a cashpoint explosion in Fürth, Hessen, around 250km from the Dutch border. Witnesses described how it was driven into Nierkerkestraat, Osdorp, at high speed, hit a speed bump and crashed. The occupants, at least one of whom was armed, got out and fled on foot.

A specialist explosives team quickly arrived on the scene and cleared all houses within a 100m radius after discovering bomb-making materials in the vehicle. Residents had to stay out of their homes for nearly nine hours following the incident, which happened at around 1pm on Tuesday, until the car was taken away on a tow truck.

Police appealed for witnesses to come forward, as well as anyone with information about the storage of explosives in the area. It is not clear how many people were in the car.

Cross-border gangs

German police said no money was stolen in the attack on the cash machine, but the building where it was located was badly damaged by the blast. The Osdorp area of Amsterdam is known to be a base for gangs that carry out cross-border raids on German cashpoints, which tend to contain more money than those in the Netherlands because cash is still widely used there. Last month seven people were injured and an entire block of flats had to be evacuated after an explosion in a basement gym next to the Osdorper Ban which was caused by home-made explosives.

Police arrested two men aged 24 and 23, both of whom were among the casualties. Prosecutors said they were suspected of planning explosions and possessing explosive materials.

This Weapon Doesn't Explode – It Shuts Down Electronics

Source: https://i-hls.com/archives/137759

Aug 07 – Modern militaries rely heavily on satellites and sophisticated electronic systems for communications, navigation, intelligence gathering and precision targeting. Rather than destroying these assets with conventional weapons, armed forces are increasingly exploring technologies capable of disabling them electronically. High-power microwave (HPM) weapons are one such approach, using intense bursts of electromagnetic energy to disrupt or permanently damage electronic circuits without relying on explosive warheads. A newly published military research paper has provided an unusual glimpse into the development of such systems, describing a new generation of high-power microwave technology capable of producing pulses of up to 100 gigawatts. The paper indicates that the technology has progressed beyond laboratory experimentation and into practical military applications.

The centerpiece of the research is a pulsed-power architecture that combines multiple synchronized microwave generators into a single system. Instead of attempting to increase the output of one large generator, which becomes increasingly difficult because of insulation and engineering limitations, the design coordinates several compact modules that fire simultaneously. By synchronizing the pulses, the system achieves a significantly higher combined output while allowing each individual module to operate at peak efficiency. According to Interesting Engineering, this modular approach addresses one of the biggest technical challenges facing high-power microwave weapons: generating extremely high energy levels while keeping the system compact enough for operational use. The paper also describes all-solid-state pulsed-power systems and hybrid lithium-ion capacitor technology capable of delivering immediate power even at temperatures as low as -40°C, potentially improving readiness in harsh operating environments.

Although the publication does not discuss specific deployment scenarios, the technology is closely tied to electronic warfare. Microwave pulses in the gigawatt range are designed to interfere with or damage electronic components rather than physically destroy a target. The researchers note that even one-gigawatt pulses can significantly disrupt low Earth orbit satellites, while substantially higher outputs could pose a greater threat to satellite constellations used for communications, navigation and military operations.

Beyond space-based targets, high-power microwave weapons are also viewed as a potential means of disabling radar

systems, communications infrastructure and other electronic equipment without kinetic strikes. Because they attack electronics rather than structures, they could offer militaries another option for suppressing enemy capabilities while limiting physical destruction.

The research also highlights ongoing efforts to improve beam control, reduce system size and lower production costs, which are areas that will likely determine how widely high-power microwave weapons are adopted in future electronic warfare and counter-space operations.

Secondary IEDs

A new article in *ASPIS Medical*, a publication of Counter-Terrorism Medicine Europe (CTM-E), examines how secondary explosive and hazardous devices are used to target police, firefighters, paramedics, bomb disposal teams and civilians after an initial attack. The piece, "*Secondary Device Recognition: Threats to First Responders in Modern Terrorism and Insurgency*," was written by retired Brigadier General Ioannis Galatas and Hellenic Police Colonel Nikolaos Rallis. It defines a secondary device as an explosive or hazardous mechanism deliberately positioned to detonate after an initial incident, using the first blast to create confusion and draw responders into a concentrated area. The authors explain that the tactic exploits predictable emergency procedures, crowd convergence and the urgency to reach casualties. Its wider purpose is not only to increase casualties but also to delay lifesaving intervention, undermine public confidence and create fear and hesitation among responders. The article reviews the use of secondary devices and "double-tap" attacks. Examples include follow-on explosives placed near evacuation routes, delayed devices intended to target investigators and bomb technicians, and subsequent strikes reported while emergency personnel were assisting victims. It also notes that chemical components have occasionally been combined with explosives, including chlorine vehicle bombs used in Iraq. However, the authors say verified examples involving biological agents are difficult to identify in the available literature. For emergency services, the article emphasizes controlled scene stabilization, layered security perimeters, visual sweeps, remote observation, bomb detection resources and managed casualty extraction. Responders must balance the urgency of medical treatment against the risk that premature entry could expose additional personnel to a follow-on attack. The authors also examine the psychological burden placed on bomb disposal personnel and other responders, including hypervigilance, decision fatigue, sleep disruption and post-traumatic stress. They argue that training, rest cycles, mental health support and reliable communications are important parts of sustaining operational performance. The article, beginning on page 41 of the July 2026 issue, concludes that secondary-device recognition is now an essential part of emergency-response doctrine and should be understood as a broader approach to operational caution, situational awareness and coordinated scene management rather than only a technical bomb-disposal skill.

Matt Seldon, BSc., is an Editorial Associate with HSToday. He has over 20 years of experience in writing, social media, and analytics. Matt has a degree in Computer Studies from the University of South Wales in the UK. His diverse work experience includes positions at the Department for Work and Pensions and various responsibilities for a wide variety of companies in the private sector. He has been writing and editing various blogs and online content for

promotional and educational purposes in his job roles since first entering the workplace. Matt has run various social media campaigns over his career on platforms including Google, Microsoft, Facebook and LinkedIn on topics surrounding promotion and education. His educational campaigns have been on topics including charity volunteering in the public sector and personal finance goals.

Soon in the Greek quiver

Infrasound As A Weapon of War

By Mustansar Siam

Source: https://www.thefridaytimes.com/12-Aug-2026/infrasound-weapon-war

Aug 12 – Victor Ofosu is a London-based international security expert, author, businessman, and independent researcher. His recent academic work challenges conventional military thinking and offers a new perspective on how sound can influence outcomes in war without relying solely on lethal force. Mustansar Siam: What inspired you to explore the use of infrasound as a weapon of war? Was there a particular event, research gap, or personal experience that motivated this study?

Victor Ofosu: The current escalation of wars around the world is concerning. I understand that due to factors such as business interest, control over global security, emotions, and a lack of adequate leadership in our respective countries, the world cannot rid itself of war. War has become and will remain a constant issue in human development. Hence, I intend to discover a framework that will allow countries to engage in wars without human casualties, displacement of people and the mass destruction of property. From my perspective, infrasound, which I believe is the future of warfare, will serve such a purpose. That is a military engagement without mass destruction; the replacement of destructive weapons with infrasound warfare.

Mustansar Siam: You challenge the traditional reliance on lethal force in warfare. How do you envision sound replacing or significantly reducing the need for destructive weapons on the battlefield?

Victor Ofosu: The world is currently experiencing a dramatic shift in warfare.

Significant innovations are taking place in military deployment; for example, a shift from deploying soldiers to the utilisation of artificial intelligence and robotics. Here, understand that although the future of warfare is becoming more strategic, sophisticated and complicated, we will also observe an unprecedented level of human casualties. I am therefore of the opinion that the development of infrasound as a weapon of war can assist in addressing the destructive nature of future escalation of wars. With the application of infrasound, we will observe zero destruction of property and human casualties. I understand that there are effects of using infrasound, such as its impact causing nausea, mood changes and depression; on the contrary, effects are less dramatic when compared with the use of nuclear weapons, guns and missiles. Most importantly, there are no casualties when employing such a strategy. I believe that if the researchers, academics and science communities embrace my research, the benefits in the future will outweigh the adverse effects of using infrasound.

Mustansar Siam: You propose a new definition of strategic communication based on psychological conditioning. What is your definition, and why do you believe psychology should be at the heart of strategic communication? Victor Ofosu: In our daily exchange of human engagement, the exchange of information is not merely a process of storytelling, but humans communicate to elicit behaviour changes in the intended audience or the receiver. Furthermore, one must understand that war is fought to subdue the will of the opponent. That is, war is not fought to tell stories, as some academics, in their disposition, define strategic communication as storytelling aimed at persuading the target audience. I intend to refute such a definition and to develop a befitting framework which provides a coherent and concise understanding of the term strategic communication. Hence, I define strategic communication "as the purposeful flow of information to stimulate, condition and change the attitude, beliefs, and behaviour of the target audience." My definition is underpinned by psychology; psychology is underpinned by the human mind, emotion, feelings and memories; this is expressed in the conditioning effect of information. Here, the role of strategic communication is to penetrate the opponent's mind and to condition him to act favourably. My definition is supported by psychology, as I argue that strategic communication effectively changes the opponent's emotions, feelings, and memories; this process activates the desired effect on the battlefield. I intend to discover a framework that will allow countries to engage in wars without human casualties, displacement of people and the mass destruction of property. Mustansar Siam: How does Pavlov's classical conditioning experiment apply to modern military strategy, especially when using infrasound as the neutral stimulus? Can you explain the mechanism you have in mind? Since infrasound is mostly inaudible, how would it be used to transmit subliminal messages or create conditioned responses in enemy forces? What role do memory and the brain's hemispheres play in this process?

Victor Ofosu: The application of infrasound is currently a concept that requires further investigation and experimentation. However, I insist that

when produced as a war machine, infrasound will perform as mentioned in my research. Subliminal information can be manufactured, as in the case of saving music on a music device. The difference here is that when the device is played below 20 Hz, making the message inaudible, it will achieve the desired effect in the targeted audience. On the battlefield, this process will be applied differently; it requires a vessel, which is the carrier, and attached to this vessel is the transmitter. The process is intended for a large population, not a small group. However, it can equally affect change in a small group. I argue that a drone can be used to carry the transmitter. When flown over the targeted location, the operator can transmit the messages below 20 Hz at the appropriate time. The effectiveness of this process is heavily reliant on Pavlov's classical conditioning, thus the utilisation of the neutral stimulus and an unconditioned stimulus. Infrasound can penetrate the skull and thus alter the functional connectivity in the brain. For example, it can alter areas in the right hemisphere of the brain, such as the amygdala and temporal gyrus. This process is significant, as the right brain controls movement, thinking, emotion, interpretation, and awareness.

Mustansar Siam: What specific effects on soldiers' morale, emotions, cognition, and combat performance do you expect from the strategic use of infrasound?

Victor Ofosu: The effect on the soldier is dependent on the response expected by the applicant. Here, it can be observed that the message delivered will activate the expected conditioned response in the combatant. Ultimately, to win a war, the applicant of this tool will expect to alter the soldier's will to fight through eliciting a negative response to combat.

Mustansar Siam: In your opinion, what will be the role of infrasound in future conflicts? Will it become a standard tool in military arsenals, or remain a niche psychological weapon?

Victor Ofosu: I intend further to examine the application of infrasound on the battlefield. Perhaps my research will assist academia to accept my findings. Naturally, I expect some academics may be critical of my observation and may argue it is impossible to apply such a method in war. However, if man can launch a satellite into space, I believe that achieving my concept is not an impossibility. I hope that soon this process will replace military arsenals on the battlefield. The application of inaudible low sound frequency is indeed the future of warfare.

ICI International CBRNE INSTITUTE

CBRNE-Terrorism Newsletter

CYBER NEWS

C2BRNE DIARY

Europol-led action against nihilistic violent extremist network "The Com"

Source: https://www.europol.europa.eu/media-press/newsroom/news/europol-led-action-against-nihilistic-violent-extremist-network-com

July 24 – Over several weeks in June and July 2026, Europol supported an action targeting nihilistic violent extremist content online. Investigators from nine countries participated in these 'Referral Action Days', with the common goal to disrupt The Com online ecosystem and limit propaganda dissemination, as well as to find new investigative leads. The action also aimed at enhancing platforms' response to and awareness of terrorist content online produced and disseminated by The Com groups.

An estimated 4 340 URLs related to nihilistic violent extremism were referred during the initiative organised by Europol's EU Internet Referral Unit - EU IRU and the Spanish Intelligence Centre against Terrorism and Organised Crime (CITCO). This successful action complements the efforts undertaken within Project COMPASS, coordinated by Europol's European Counter Terrorism Centre, which unites law enforcement authorities from EU Member States and third-party partners to address The Com.

What is 'The Com'?

The Com, referring to the "Community," is a loose network of fringe online groups that adhere to a misanthropic and nihilistic worldview. These groups lack a unified ideology or centralised structure, and within this ecosystem, violence is often an end in itself. Some factions have adopted violent right-wing extremist beliefs and accelerationist views, aiming to hasten societal collapse through violent attacks and the corruption of youth.

Groups affiliated with The Com recruit members and groom victims on social media, messaging apps, and gaming platforms to engage in self-harm, animal torture, violent attacks, and the production of child sexual abuse material. They distribute propaganda on accessible platforms to attract young individuals, funnelling potential members and victims into private forums and chat rooms where radicalisation and victimisation occur. Victims are typically coerced into remaining under the perpetrators' influence through (s)extortion.

While the use of widely accessible social media platforms and online games to recruit and to target victims is not new, the ease of access to The Com-related content is alarming. Social media algorithms expose young individuals to extremist material - whether they seek it or not - with content that may initially appear harmless but quickly escalates into extreme violence and traumatising material. The Com's content is laced with coded language and emojis that adults may struggle to decipher, conveying sinister messages that encourage self-harm and the sexual exploitation of minors.

Horrific online content disseminated by The Com

Within The Com, content is used to promote an online alias or a group's notoriety. The more extreme and harmful the content a user or group can produce or extort, the higher their status within the online community. These acts are often livestreamed on social media platforms, where online bystanders cheer them on, and later saved and disseminated. The content referenced in this operation includes violent videos and images depicting self-harm, suicide, child sexual abuse material (CSAM), animal cruelty, and violent attacks.

Within The Com, so-called blood walls and cut-signs are promoted and sometimes required as entry criteria for certain groups. Blood walls are inscriptions or paintings made with blood that display the extorter's online alias and group affiliation. The Com frequently employs satanic expressions and symbols to make its content appear more sinister and frightening. Some groups also incorporate violent right-wing symbols and jargon, either to convey their ideological stance or to shock audiences. Cut-signs involve individuals carving the extorter's name into their own bodies, with images or videos of these acts later shared to enhance the extorter's and their group's status within The Com.

Other types of content include videos of violent street attacks, a practice known as "manhunt," or arson. Additionally, The Com produces manuals or guides covering topics such as grooming, murder, improvised explosives, or ideological viewpoints. These manuals are created by groups to instruct members on methods for committing violent attacks, grooming and (s)extorting vulnerable minors, or conducting doxing and swatting.

Doxing involves uncovering and exposing an individual's personal data to extort or harass them. Personal information, such as a home address, may be used to carry out "swatting". This is a tactic where the extorter contacts law enforcement with a hoax emergency call or threat, falsely claiming a violent crime is about to occur at the victim's address. This prompts an emergency response, often endangering the individual and their family. The procedure is used to intimidate victims and may cause physical or psychological harm.

Europol's support

Europol provided comprehensive support for the Referral Action Days (RAD) by coordinating operational meetings with

the participating countries, delivering presentations on the phenomenon, and collecting and sending a consolidated list of contributed URLs for deconfliction. Additionally, Europol collected and referred content that violates the EU Directive on Combating Terrorism and the EU Framework Decision on Combating Racism and Xenophobia, while supporting third parties in their referrals. Europol produces a strategic analysis report of the action to enhance understanding of the phenomenon and improve response efforts.

'The Com' in the context of the EU terrorism landscape

In the recently published European Union Terrorism Situation and Trend Report (EU TE-SAT 2026), nihilistic violent extremism is identified as a destabilising ideology, with "some subgroups and individuals within this ecosystem explicitly seeking to undermine societal structures." The focus of the Referral Action Days stems from law enforcement authorities' need to address nihilistic violent extremist online venues, which are misused to groom, radicalise, and disseminate terrorist content online. Over the past two years, Europol's European Counter Terrorism Centre has observed The Com network evolving into a global threat, particularly concerning minors as both victims and perpetrators. The ECTC has received hundreds of requests from Member States and third parties regarding crimes committed within The Com milieu. Requests related to nihilistic violent extremism have also been reflected in investigations submitted to the ECTC for support.

A unified response to terrorism and violent extremism

The joint action ties into the European Commission's counterterrorism agenda ProtectEU. One of its central pillars is protecting people online, and the Referral Action Days specifically contributed to Europol's action points on facilitating EU-level cooperation with online service providers. This includes improving the handling of illegal content related to terrorism and other harmful material, as well as monitoring extremist misuse of gaming platforms and providing regular assessments.

Participating countries

Belgium, Finland, Hungary, Ireland, Luxembourg, Netherlands, Portugal, Spain, Sweden.

What Happens if China Hacks the US Water Supply? I Went to a Secret War Game to Find Out

By Andy Greenberg

Source: https://www.wired.com/story/what-happens-if-china-hacks-the-us-water-supply-war-game-volt-typhoon/

It's around an hour and 10 minutes into the role-playing game I've been invited to observe, a simulated catastrophic cyberattack on US water utilities, when the whole thing begins to feel less like a fun afternoon playing Dungeons & Dragons and more like a plausible threat to civilization.

A full 24 hours of in-game time have passed since hackers disrupted 5,000 water utilities across the United States in this imagined scenario. Joshua Corman, the former Cybersecurity and Infrastructure Security Agency strategist serving as our dungeon master, stands at the front of a conference space in an office tower high above Times Square, narrating the latest updates to the game's participants, a few dozen insurance executives set up in six teams. All of them have gone disturbingly silent.

"You ready? It's about to get harder," Corman says. "I'm going to share a few things, and it's going to hurt."

It is, of course, still the same April afternoon as when we started—but in game time, the second-order effects of widespread water outages have started to become clear. Food refrigeration systems are failing at cold storage warehouses. Water-dependent drug and chemical manufacturing has been bottlenecked, leading to insulin shortages. Data centers' cooling systems are failing, causing outages of cloud services. Most critically, 2,000 hospitals are without water, hampering patient care and in some cases leading to evacuations as HVAC systems shut down and the July heat—the game takes place just before Independence Day in 2027—bakes facilities.Worse yet, Corman is playing a looping video onscreen, at the front of the room, showing a burst water main: The hackers have managed to trigger not just IT disruption but also, in at least some cases, real physical destruction that will take far longer to fix. "Everyone downstream is without water pressure," Corman says. "Everything depends on water." (The tension in the room has also peaked, in part, due to Corman's decision to deny participants any organized restroom breaks. "There are no breaks in real incident response," Corman explains just before the giant water pipe starts gushing onscreen. "If you have to go to the bathroom, go to the bathroom. But you might miss something vital." No one goes to the bathroom.)

The central task Corman has assigned for this round to the teams of participants, all of whom are playing the surprisingly pivotal role of insurance companies, is to decide how they'll dole out their resources—contracted cybersecurity incident responders and money—and which clients will get priority.

Will their business relationships with clients dictate their response? Or will they focus on minimizing harm for the most possible people? And given that some clues in the game are already suggesting this catastrophic cyberattack was carried out by the Chinese military to hamper a US response to its invasion of Taiwan, will the insurers be required to focus on keeping military facilities operational? Left unspoken in this question is another range of disastrous possibilities for the insurers themselves: Will this catastrophe bankrupt them? Or will they invoke an "act of war" exclusion in their insurance policies—a standard clause that exempts carriers from all liability when an armed conflict breaks out—and pay their clients nothing, thus risking that they'll become the villains of the story? The teams have 15 minutes to decide on a policy. "OK?" Corman asks. "Let's start the clock now."

Most China-watchers in the cybersecurity world agree, in fact, that this particular clock has already been ticking for years.

In May 2023, Microsoft, the National Security Agency, and CISA all announced the discovery of what they called Volt Typhoon, a group of hackers working in service of the Chinese military. The intruders had broken into the networks of critical infrastructure facilities across the continental United States and the US territory of Guam, hitting targets related to everything from manufacturing to telecommunications to the electric grid. These breaches were especially alarming because the hackers seemed to be going beyond the espionage that's become standard practice for Chinese state cyberspies. Instead, according to Microsoft, they were "pursuing development of capabilities that could disrupt critical communications infrastructure between the United States and Asia region during future crises." Volt Typhoon was, in other words, "pre-positioning," as another CISA and NSA advisory would put it in early 2024, laying the groundwork for broad cyberattacks aimed at disabling the US military at a crucial strategic moment—perhaps, some cybersecurity analysts suggested, on the eve of an invasion of Taiwan.

As the US government and cybersecurity industry continued to track Volt Typhoon, however, it became clear that the hackers' target list wasn't limited to networks that would allow the sabotage of US military assets. They included the IT systems of a water utility in Hawaii, multiple US ports, and at least one oil and gas pipeline that might have military relevance, but also hundreds of other entities including water and electric infrastructure as small-scale as the Littleton Electric Light & Water Departments in Littleton, Massachusetts, a town with just under 10,500 residents. "The only reason to target that sort of entity is to cause societal chaos in the United States," CISA's former executive director Brandon Wales told WIRED in early 2025. Volt Typhoon, he said, seemed to be preparing to "cause chaos in the homeland—to influence our geopolitical freedom of action, our willingness to fight."

Even now, three years after Volt Typhoon's initial discovery, threat intelligence analysts say China's efforts to prepare for US civilian infrastructure disruption continue. Joe Slowik, a former Los Alamos National Labs cybersecurity researcher working on contract for the Department of Energy, says Volt Typhoon—or a related hacker group it has evolved into—is still targeting the US electric grid and water utilities.

Some of these intrusions are caught, Slowik says. Others go undetected, in part due to the minimal security budgets of municipal utilities and in part due to the hackers' stealthy mode of operating, known as "living off the land," that hijacks legitimate functions in a network instead of planting malware. "It's pretty good tradecraft," says Slowik, who now leads threat research at the cybersecurity firm

Dataminr. "But it's also applying that tradecraft to areas that really don't have the capacity to identify it."

The scenario modeled by Corman's war game—5,000 hacked water utilities—would be unprecedented. It also isn't the most probable outcome of Volt Typhoon's intrusions, cautions Jen Easterly, who served as director of the Cybersecurity and Infrastructure Security Agency when China's hacking campaign was first discovered. Yet Easterly, now CEO of the RSA cybersecurity conference, also warns that AI could make that mass-sabotage hypothetical far more plausible, particularly over the next few years, when its use in offensive hacking may outpace its use by defenders.

The scale of China's preparations remains unknown, Easterly says, but its intent is clear. "What we found was really just the tip of the iceberg," she says of her time running the Volt Typhoon response at CISA. "Do I believe there's any change to China's very deliberate strategy to create access points in our most important civilian infrastructure, to be able to launch disruptive attacks in the event of a crisis in the Taiwan Strait? No, I don't."

Just two months ago, former NSA director of cybersecurity Rob Joyce spelled out the warning even more starkly in an article for the Cyber Defense Review, arguing that the threat Volt Typhoon poses remains as clear and present as ever.

"China has effectively strapped the digital equivalent of explosives to the backbone of American society," Joyce wrote. "Quietly maintaining access. Waiting."

Even as it has loomed over US cybersecurity for more than three years, Volt Typhoon—or whatever it has now become—has never in a single confirmed incident pulled the trigger and carried out an actual disruptive cyberattack. Understanding the threat it represents remains an exercise in imagination.

Hence the war game that I'm here to witness. The event, convened by an insurance industry cybersecurity organization called CyberAcuView, has invited 30 or so insurance execs. They've allowed me to sit in on the condition that none of them or their employers are identified.

Corman, who has run dozens of these types of war games, says he asked me to attend this specific exercise for two reasons: First, few people understand the central role of insurance in the midst of a cybersecurity emergency. Hacking victims' first call is often to their insurance company, who then approve and unlock law firms and cybersecurity incident response providers in the hours that follow.

Second, Corman argues that insurance companies serve as illuminating game players: They have a strong financial incentive to squarely assess risk—"skin in the game," as Corman puts it—without hyping it up or downplaying it.

Around 1 pm, after some introductions and ground rules ("Don't fight the scenario"), Corman begins. He quotes the 1983 hacker classic film *WarGames*, addressing the room: "Shall we play a game?"

Day one is July 1, 2027. The New York Times has published a headline that morning about growing fears of a Chinese invasion of Taiwan. A cybercriminal ransomware gang's ongoing hacking spree has taken up about a third of the US's cybersecurity incident response capacity. The cybersecurity industry is warning, meanwhile, of three vulnerabilities in network edge devices like firewalls and VPNs—the kind that Volt Typhoon has long used to get a foothold on a victim network—being exploited by unnamed hackers.

Then the participants hear the real starting gun: A restricted advisory from the US federal government states that thousands of water utilities have been breached, their controls are unresponsive, and anecdotal reports suggest physical damage. The execs' first assignment is to determine if and how they'll communicate this news to customers, and then—given that this will almost certainly exceed their capacity to unlock funds for targeted victims—decide how they'll choose which clients to prioritize.

The breakout sessions start. At my table, an argument begins almost immediately on the question of whether to alert the insurer's broad range of corporate clients. Most of the table wants to share the news with all customers. One player disagrees, arguing that the usual insurance model is that it's the client who first reports an incident to the insurer along with a claim—not the other way around.

"So do? Don't?" one player asks the table.

"Don't," the holdout says. The naysayer is quickly outvoted, and they choose to alert all their customers.

Corman is walking around the tables, handing out surprises. He comes to my table, tells someone to roll a 20-sided die, then looks at the result and tells the group without explanation that all incident responders at Dragos, CrowdStrike, and Mandiant—some of the biggest providers of those emergency cybersecurity responses for industrial services like water utilities—are unavailable, apparently busy with other victims. My table will have to scrounge up incident responders from other, smaller companies or beg for the help of government agencies. "Perfect," one player responds.

This scarcity makes the question of which customers to help first all the more touchy. The table comes to a rough consensus that they'll prioritize clients in order of size, biggest first, defined by revenue.

As each table takes turns presenting its decisions, other tables say they've opted to prioritize customers on a first-come, first-serve basis, or to follow some vaguely defined sense of "national security" as defined by the government, though it's not clear who in the government is going to offer that definition.

One table asks where incident responders from CISA, the nation's premiere cybersecurity defense agency, have been during all of this. Corman notes that CISA has spent more than a year without a Senate-confirmed director and has hemorrhaged staff. Some tables suggest they'll ask for help from cybersecurity experts in academia or the National Guard.

None of that, it's soon clear, will be enough.

Day two of the simulation arrives, and Corman lays out the new reality: Numerous water mains across the country have broken. The man-made drought has spread to hospitals, data centers, refrigeration, and manufacturing.

Then Corman throws another curveball: He plays a prerecorded video statement from a fictional military official appealing to the insurance companies' help in responding to the geopolitical threat posed by China, the first time that country's name has been spoken in the game so far. "I'm most concerned about our ability to protect our military mobility, a key element of national security," the official tells them.

Corman hands out the day-two assignment: As the disruption spirals outward, how will they prioritize *now* which of the water utilities deserve their resources? The "biggest-customers-first" or "first-come, first-serve" answers from the prior round, just a few minutes earlier, now seem hopelessly naive. Will they focus on restoring water in places where they can save the most lives, such as hospital-dense cities? Or will they seek to minimize economic harm? Or heed the military's request to focus on national security, essentially prioritizing the military's response to China's potential invasion of Taiwan?

Fortunately, no one in the room is a monster. After 15 minutes of breakout conversations, the teams around the room render the same verdict, that their first priority will be to save human lives—though none spells out how they'll make the endless impossible decisions that follow from that answer.

Only one person, after all six teams have given that same answer, speaks up to raise an uncomfortable point. Prioritizing harm to people above all else may not be an option. "The easy answer is public safety, human life," he says. "The more difficult one is when you do have regulators or someone calling, shareholders asking questions."

"If Treasury is calling and asking numbers, and we're saying we're focused on human life, I don't know if that's the actual talk track," he goes on, using a sales term for a phone script of talking points for client conversations. Or, he adds, if an official is telling the company it needs to focus on telecommunications or "dual use" infrastructure—meaning things that might have military importance—that might become "priority number one," he says.

In other words, taking the most direct action to protect people from harm in the midst of a catastrophic cyberattack might require breaking contracts, flouting the military's demands, or directly contradicting a larger US government strategy in the early days of an unfolding war.

"We didn't agree on that as a table," he says. "There's not going to be a consensus."

At this point, abruptly and mercifully, Corman ends the game to start a lessons-learned session. During this round, he's put up a slide that represents some of the infrastructure disrupted by the second-order effects of the hackers' cyberattacks. Next to each is a long line of multicolored dollar signs and outlines of people, representing financial loss and human casualties.

There's no point in counting these as if they're some sort of score or demerit, Corman assures me when I ask afterward. They're less a quantified measure of losses than a qualitative assurance that things have gotten very bad. He's gotten his point across: If this game has any winners, they're not in the room.

In a postmortem call later on, I ask Corman: Was it ever actually possible to win his game? Or even to change its outcome? Or was it a kind of *Kobayashi Maru*, an impossible test designed to teach some other sort of lesson?

Corman gets the *Star Trek* reference but disagrees with the analogy. There were moves the players could have made that mattered, he says, such as prioritizing water recovery in hospital towns from the start. He was also interested to see if any of the teams invoked exclusions to their insurance policies based on acts of war—an argument that led to drawn-out legal battles following the billions of dollars in damage from Russia's NotPetya cyberattack in 2017—or planned to claim reimbursement from the federal government for their costs under the Terrorism Risk Insurance Act, or TRIA. None of that came up in my table's game, however, due in part to the lack of certainty on day two of who was behind the hacking campaign.

But yes, Corman adds: The goal of the game was not to create a competition to be won. Instead, it was designed to "overwhelm"—to "surface and shatter assumptions."

More specifically, Corman says, he wants to show the industry that if even a relatively mild version of Volt Typhoon's potential hacking occurred under current conditions—Corman argues the real version could be far *worse* than 5,000 disabled water utilities—it would be beyond the insurance industry's capacity to handle it.

Mark Camillo, CyberAcuView's CEO, who cohosted the war game with Corman, says the results show that a cataclysmic cyberattack may well be "uninsurable": that the costs of such an event would surely bankrupt the industry if insurers don't invoke act-of-war exclusions that would get them off the hook—which would in turn cause enormous damage to the trust that American society puts in insurance. One lesson of the game, Camillo contends, is that insurance companies might need a government fund like the terrorism-focused TRIA to help cover its costs, to be refilled with payments from policy holders in the years following a cyber disaster.

But Corman says the bigger lesson is that the insurance industry, maybe even more than the US government, has the unique power to change all of this—not by learning how to respond in the midst of a hacking catastrophe but by focusing on how to prevent one. It could, for instance, demand via insurance policies that customers better protect themselves, compelling clients to review their networks for the unpatched, vulnerable network devices that Volt Typhoon exploits, or requiring that they join cybersecurity information sharing groups. Currently, Corman says, just 0.3

percent of the country's 151,000 water utilities are part of those organizations, an appallingly low membership compared to similar groups in industries like finance or electrical utilities. "The idea is to help insurers realize that this is not going to play out the way they thought it would," Corman says in summary. "And make them think about what they might do differently once they leave the room."

As the final lesson from *WarGames* goes, after all, there are games where the only winning move is not to play—or at least not on the adversary's terms.

Andy Greenberg is a senior writer for WIRED covering hacking, cybersecurity, and surveillance. He's the author of the books *Tracers in the Dark: The Global Hunt for the Crime Lords of Cryptocurrency* and *Sandworm: A New Era of Cyberwar and the Hunt for the Kremlin's Most Dangerous Hackers*. His books and excerpts from them published in WIRED have won awards including two Gerald Loeb Awards for distinguished business and financial reporting, a Sigma Delta Chi Award from the Society of Professional Journalists, and the Cornelius Ryan Citation for Excellence from the Overseas Press Club. Greenberg also hosts WIRED's video series Hacklab, which was a finalist for a National Magazine Award.

This Coin-Sized Device Can Hack a Boeing 737

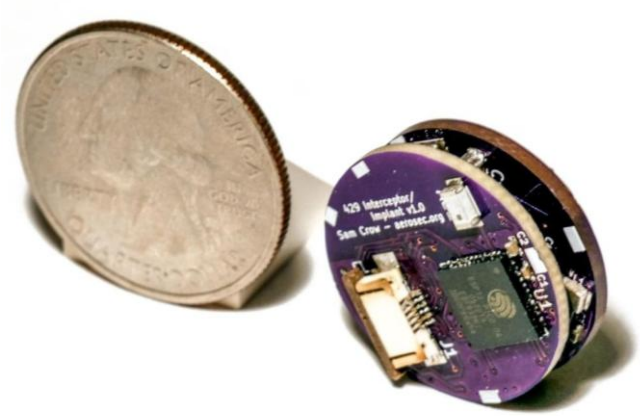
Source: https://www.wired.com/story/this-coin-sized-device-can-hack-a-boeing-737/

Aug 12 – Even as the digital components of so many life-critical systems have proven susceptible to cybersabotage—cars, medical devices, even water utilities and power grids—the computer systems of airplanes have, thankfully, remained uniquely inaccessible to hackers. But one group of academic researchers has spent years testing a different, devious approach to aviation cybersecurity. Perhaps, they suggest, a plane could be hacked the same way that spies and saboteurs have targeted other high-value, offline computers: by surreptitiously gaining physical access to one and plugging in a device designed to silently run the attackers' malicious code.

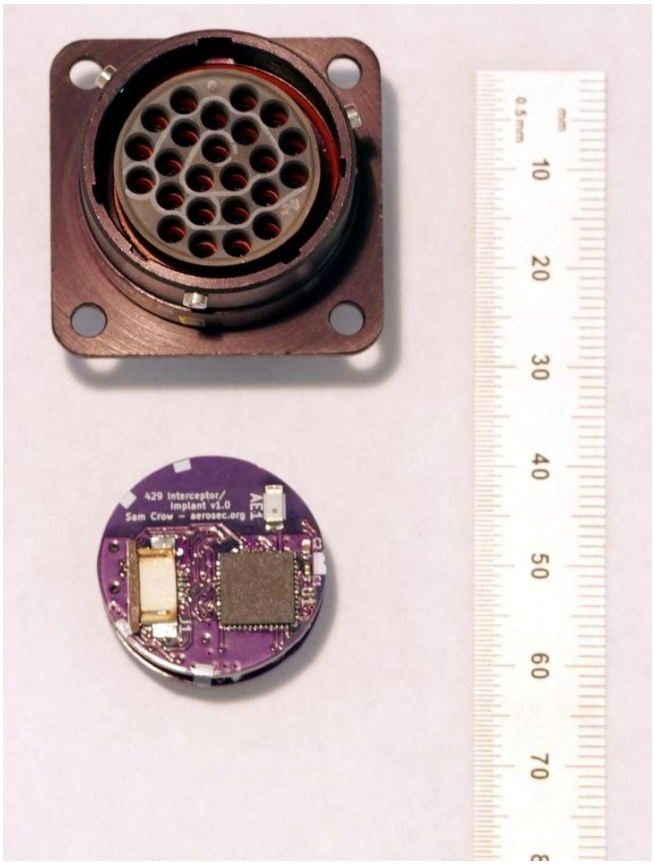
Tomorrow at the Usenix Cybersecurity Conference, researchers from the University of California at San Diego and Oberlin College will present a hacking technique capable of commandeering the autopilot of a Boeing 737 to redirect its navigation or silently altering key values in the plane's takeoff and fuel calculations while spoofing the results on the pilot's screen—subtle changes the researchers say could potentially cause anything from runway overruns on takeoff to diversions to a different country's airspace to catastrophic crashes.

To carry out that hacking, they've built a roughly coin-sized, Wi-Fi-enabled prototype device that costs less than $100. In less than a minute, that hardware implant can be fitted into a port accessible via a hatch on the exterior of the plane, one that's routinely within reach of maintenance workers or other airport and airline staff between flights. Once it's in place, the device can send electrical signals on one of the 737's internal networks to spoof commands to sensitive computer systems that guide its autopilot and show the pilot

variables like the plane's total weight and outside air temperature, which play a critical role in a 737's takeoff calculations.



The researchers' hacking device, next to a quarter for scale. | Photograph Courtesy of UCSD



The hacking device (bottom) next to the connector it fits into on a Boeing 737, one that can be found inside a port that's accessible from a hatch on the plane's exterior. | Photograph Courtesy of UCSD

By proving the viability of that technique, the result of a process that stretched over more than a decade and entailed buying tens of thousands of dollars' worth of plane components for testing, they hope to show that this sort of physical access hacking represents a practical threat in the hands of well-resourced saboteurs and a significant blind spot in aircraft security. Compared to the traditional threat of simply planting a bomb on a plane, they argue, it's also an approach that would offer an attacker more control, stealth, and deniability.

"If you could get 60 seconds with an airplane, what could you do?" asks Stefan Savage, one of the UCSD computer science professors who led the project, describing the question that first motivated their line of research. "Well, it turns out there's a port that's externally accessible. You can get to it with no special tools in about 15 seconds. And you can shove in a piece of electronics a little bigger than a quarter that lets you basically tell the autopilot what to do and lie to the pilot about changes to the flight plan."

The researchers aren't revealing which port they targeted on the 737, nor are they releasing some details of how their hacking device is able to spoof commands to the plane's computers. They've worked closely with Boeing to share their findings, first disclosing elements of their research to the company more than six years ago, and going so far as to test out and demonstrate their attack in a Boeing facility's test lab. When WIRED reached out to Boeing about the researchers' work, it responded in a statement that it had carried out its own review of its components' designs, installations, and interfaces in response to the researchers' findings. But it downplayed the practical risk of their physical-access hacking technique. "Our technical experts are confident that the layers of protection in place on the airplane, including within the system design and the operating environment, provide sufficient mitigation to significantly limit the feasibility and risk of real-world attacks," the statement reads.

For their part, the researchers say, Boeing hasn't told them about any technical fix for the vulnerabilities they've discovered—and they speculate that the company may not in fact implement any such update to their systems for years to come, given how rarely commercial airplanes are redesigned. That lack of an immediate security update for planes shouldn't be cause for panic or grounding aircraft, they write in their paper. "All of the authors of this paper routinely travel on Boeing 737 aircraft and expect to continue doing so," the introduction of the paper reads.

Savage argues, though, that the research has demonstrated the need for long-term changes in both the cybersecurity of airplane components and, perhaps more immediately, the operational security measures that determine who can access a plane while it's on the ground. Their simplest fix suggestion: Plug the port with epoxy, or remove it altogether. "This is something the aviation industry will want to plan to defend against," Savage says. "I would not sleep on this one."

Building a Plane, Then Breaking It

This particular team of researchers' interest in hacking a plane originated

nearly a decade and a half ago, when some of them discovered and demonstrated the first successful over-the-internet techniques for hacking a car's computer systems, including its steering and brakes. Their proof-of-concept attack methods, particularly ones carried out by exploiting a Chevy Impala's OnStar system, launched an era of automotive hacking research that ultimately led to a sea change in carmakers' cybersecurity practices, including launching bug bounty programs for cars and hiring car hackers to help them root out vulnerabilities.

In the wake of that car-hacking work, one member of the team, then UCSD research scientist Kirill Levchenko, suggested they try hacking airplanes next. But unlike a Chevy Impala, a Boeing 737 was well beyond their budget. "I pointed out that we can't exactly buy a plane and put it in the parking lot, but he was undeterred," Savage says.

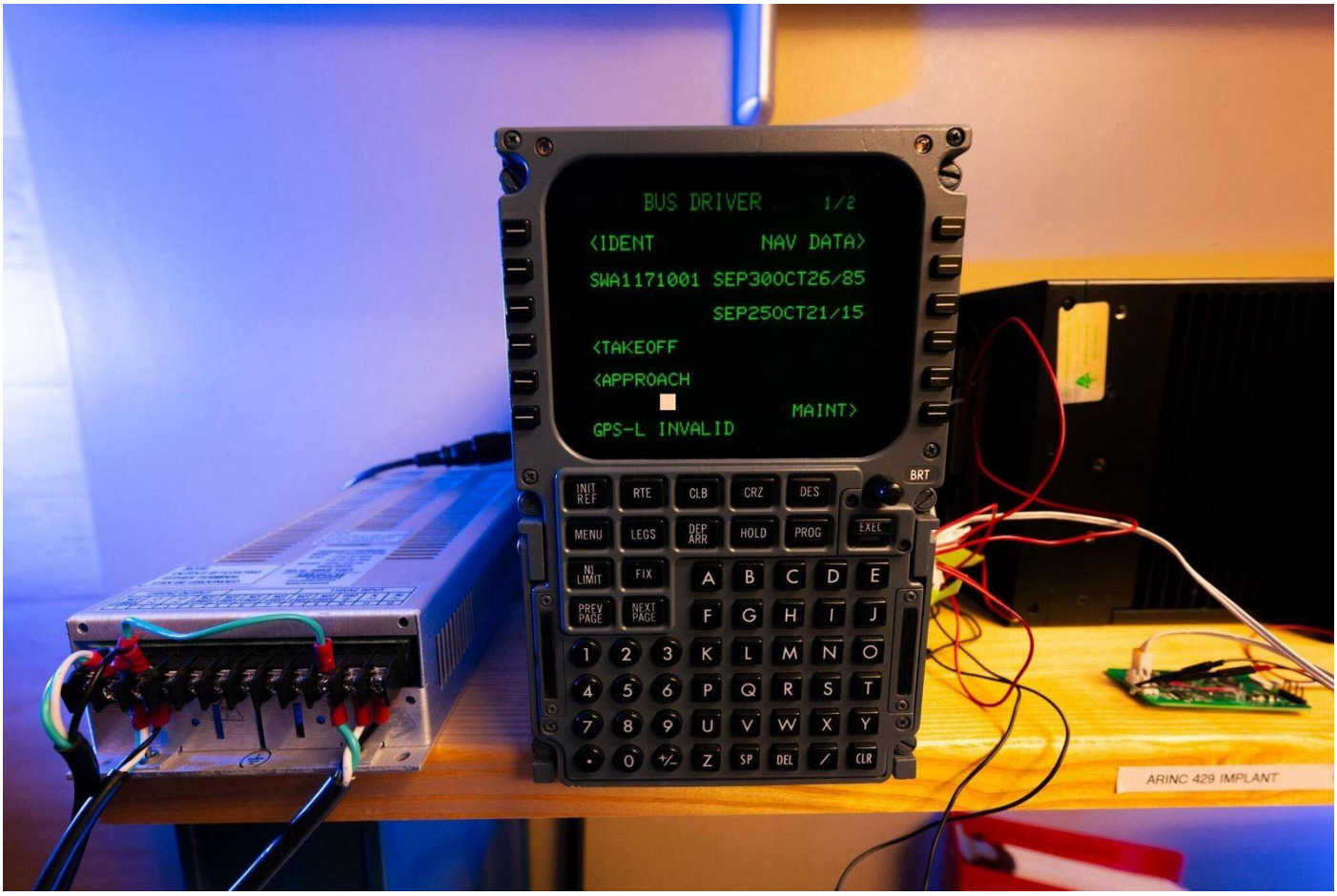
UCSD professor Aaron Schulman pushing a button on the Flight Management Computer that's part of the researchers' Triton avionics test bed, a collection of wired-together Boeing 737 computer components. | Photograph: Erik Jepsen/University of California San Diego

Over the following years, the team began buying computer components from that commercial aircraft whenever they could find them for sale, spending tens of thousands of dollars to acquire the equipment on the secondhand market. By 2019 they had assembled what they called Triton, an "avionics test bed" that essentially consisted of wired-together 737 computer parts.

Around the same time, UCSD professor Aaron Schulman was working on another research project on credit card skimmer devices that hackers were physically planting on gas station point-of-sale terminals to steal payment information. "We realized that it's a reasonable threat for someone to plug a device into a bus and read stuff off of it and potentially even gain control of it," says Schulman, using the term "bus" to mean an internal network connecting components of a piece of digital equipment. Schulman began to wonder if a similar physical access attack might work on the internal bus of a plane. "We were like, 'Wait a minute, we've got to rethink everything.'"

With that idea in mind, one of Schulman's then student researchers, Sam Crow, hunted through hundreds of pages of Boeing wiring diagrams and found that one particular port—typically protected by only a hatch without a lock—connected to a certain 737 bus that carries the data for two crucial computer components on the plane, its Flight Management Computer and its Multipurpose Control Display Unit. Soon after, Crow discovered, using the group's avionics test bed, that when he connected to that bus and sent electrical signals with a higher current than the legitimate ones

used to send commands between those components of the 737, he could override those commands with his own, a technique that the researchers would later dub “Bus Driver.” Now he had found an accessible foothold on the network from which he could send those electrical commands and tamper with the plane’s flight plan and critical parts of the pilot's interface. As Savage puts it, “it was like he had found the goddamn exhaust port on the Death Star.”



The Flight Management Computer, altered to display the name of the researchers' hacking technique, Bus Driver. | Photograph: Erik Jepsen/University of California San Diego

A Tiny Stowaway With In-Flight Wi-Fi

In the spring of 2020, the team alerted Boeing to its findings and got a surprisingly interested response. They would continue to share updates with the company on their findings for years to come.

Crow quickly refined their hacking device until he had a version of it that was small enough to fit into the exposed port on the 737 under a dust cap that typically covered it, entirely hiding the device from view. The tiny gadget included a chip capable of running their attack code as well as a Wi-Fi radio. That radio would, in theory, allow the hacking device to connect to the internet via the plane's in-flight Wi-Fi network and then beacon out to whoever controlled it, allowing the attacker to remotely control the device and the commands it sent. When Covid hit, Schulman shipped the team’s avionics equipment to Crow’s Bay Area home, and he continued working on it in his bedroom. Soon he had improved their device's ability to imperceptibly intercept, alter, and spoof commands to the plane's sensitive systems: It could, for instance, connect to the Flight Management Computer and alter the waypoints programmed into the plane's autopilot, redirecting the plane's flight, or alter variables like its weight or the outside temperature. By electronically tampering with the Multipurpose Control Display Unit, meanwhile, it could prevent those changes from showing up on the pilot's screen.

Trick the pilot into thinking the outside air was colder—or the plane's load of passengers and cargo was lighter—than in reality, and the 737 might not achieve the necessary speed for takeoff before running out of runway. Mess with the flight plan, and you could cause the autopilot to change the plane's heading to make it enter another country's airspace, where it could be commandeered by that country's air force. A sudden navigation change could potentially crash a plane into a mountain, or a slow one could send a transoceanic flight in the wrong direction until it ran out of fuel over water. “It could be something as subtle as, you're in the Pacific, you see blue everywhere, and this diverts you 3 degrees off course, and now you're in the middle of nowhere,”

Schulman says. The researchers note that a careful pilot would be able to recover from almost any of the attacks they've imagined: Taking manual control of the plane overrides its autopilot, and even if the Multipurpose Control Display Unit were hacked, the correct values would show up on a different screen in the cockpit. But even in this scenario, Schulman says, the pilot "would see that this is not lining up, but they would have no idea why, and it would be very confusing and probably lead them toward an uncertain conclusion about what to do next." In a less optimistic scenario—or if the hacker implements a more subtle change—Schulman says a pilot might not notice until it was too late. In their paper, the researchers outline a range of fixes for the vulnerability they've uncovered, starting with removing the connector in the vulnerable port altogether, or plugging it with epoxy. More long-term, though, they suggest planes' systems could be updated to include defenses in their software that detect their Bus Driver hacking technique, or that better electrically isolate systems, as in some military aircraft, or even add cryptographic authentication to prevent spoofing of signals among the plane's systems. Calling for these kinds of updates—not just for Boeing, but across the aviation industry—is far from alarmist given the practicality of the attack the researchers describe, says Beau Woods, a cybersecurity consultant who has served as an adviser to the Cybersecurity and Infrastructure Security Agency and as a member of Boeing's Industry Cyber Technical Council. "It is entirely possible to have someone who is on staff go up to an airplane when it's on the ground, going through maintenance, and put this type of thing in there," says Woods, who read the researchers' work ahead of publication. The paper, he says, "looks like solid empirical evidence about some realistic scenarios for high-capability adversaries."

The researchers' technique, he says, shows how the "threat model" for any highly sensitive system has to change as potential attackers' technology advances—in this case, as it became possible to fit an entire hardware setup capable of connecting to a plane's Wi-Fi and relaying commands to its systems onto a tiny disc hidden inside the dust cap of an obscure plug.

"Now that the research has been published, it can be understood and recognized that the reality has changed," Woods says. "Threat models from the 20th century rarely survive contact with 21st-century tools and techniques."

Iran-linked hackers blamed for cyber-attack that shut down UK power plant

Source: https://www.theguardian.com/world/2026/aug/23/iran-linked-hackers-blamed-cyber-attack-british-power-plant

Aug 23 – Hackers linked to Iran have been blamed for a cyber-attack that caused a British power plant to be temporarily shut down.

The incident involved a small-scale energy generator, according to the UK government, which said that at no point was there a risk to the wider energy system.

However, it marks an apparent escalation in the threat posed by Iran after the UK said it had given permission for the US to launch "defensive" operations against Tehran from British bases.

The power plant was shut down for four days as a result of the attack last month,

according to the Sunday Telegraph, which first reported it.
A spokesperson for the Department for Energy Security and Net Zero said: "This story refers to an incident impacting a small-scale energy generator, and at no point was there a risk to the wider energy system. The UK has a highly resilient energy system. We work closely with the energy sector to protect infrastructure and ensure the highest security standards." The National Cyber Security Centre (NCSC), which deals with attacks on critical infrastructure, is understood not to have received any reported outages from regulated operators of power stations.
The Conservatives said the incident reflected a "new kind of warfare", which meant Britain needed to prioritise "cheap, reliable energy". "The world is getting more dangerous, which is why we need to prioritise our energy security," said Claire Coutinho, the Conservative spokesperson on energy.
"Leaving ourselves reliant on imports of gas and electricity and failing to build enough reliable gas or nuclear plants for backup at home, all leaves us more vulnerable."
Hostile states such as Russia, China and Iran are increasingly targeting systems behind the UK's key services, according to a warning this year from Richard Horne, the NCSC's chief executive.
The government said last month the UK was "ready to defend itself" after Iran's military warned any bases being used by the US were "legitimate targets".
The UK has allowed the US to launch so-called defensive operations from British bases hosting American planes since the start of its war on Iran but has refused to help in offensive operations. That policy has not been altered by the new prime minister, Andy Burnham, who was notified last week that a decision had been made to extend the agreement with the US.
Iran's Islamic Revolutionary Guard Corps (IRGC) said last month that "any base used for aggression against Iranian territory constitutes a legitimate target for our forces".
Iran has been accused for years of carrying out cyber-attacks on various countries, including in relation to a massive power outage in Turkey in 2015 and several possible breaches of Israeli government websites in 2022.
US government security agencies issued a warning earlier this year of cyber-attacks on critical infrastructure by hackers linked to the IRGC.
The US has alleged that an Iran-affiliated group known as "CyberAv3ngers" carried out a campaign against it in 2023 that compromised at least 75 devices in multiple infrastructure sectors.

ICI International CBRNE INSTITUTE

CBRNE-Terrorism Newsletter

C²BRNE DIARY

& Robotic

DRONE NEWS

This is how China sees future wars ...

Meet the Robot Dog Built to Survive Arctic Ice and Freezing Water

Source: https://i-hls.com/archives/137218

July 30 – Collecting data in the Arctic is often as challenging as the research itself. Beneath seemingly solid snow can lie hidden pools of icy water, unstable ice layers, and terrain that changes constantly with weather conditions. These hazards make field operations expensive, slow, and potentially dangerous for scientists and rescue personnel working in remote polar regions.

Researchers recently tested a new approach: sending a compact autonomous quadruped robot into environments that humans would rather avoid. During an Arctic expedition, a modified version of the Lynx S10 robot successfully traversed Arctic Ocean ice floes, becoming the first four-legged robotic platform reported to operate in these conditions.

The robot was designed around autonomous navigation and terrain adaptation. Using four high-dynamic-range cameras and front-and-rear LiDAR sensors, it continuously builds a three-dimensional map of its surroundings. This allows it to identify obstacles, estimate safe routes, and navigate without requiring constant operator control.

The standard platform already offers considerable mobility. Weighing less than 20 kilograms, it can be carried and deployed by a single person. Sixteen precision joints allow it to walk, climb over obstacles up to 50 centimeters high, transition between wheeled and legged movement, and even stand upright temporarily to improve visibility.

Arctic conditions, however, required significant modifications. According to Interesting Engineering, engineers replaced the robot's standard wheels with specially designed biomimetic paws inspired by polar bear feet. These enlarged feet distribute weight across a larger surface area, reducing the likelihood of sinking into soft snow. Anti-slip textures and integrated crampons improve traction on ice, while upgraded sealing protects the system from water intrusion.

Researchers also adapted the limbs to perform better in mixed ice-and-water environments. By increasing the effective surface area of the legs, the robot gained a paddling-like capability that helped it move through slushy terrain.

During testing, the robot encountered snow-covered areas concealing meltwater beneath the surface. Despite these challenging conditions, it maintained stability and continued navigating through environments that would be difficult for many conventional robotic platforms.

From a defense and security perspective, robots capable of operating autonomously in remote and hazardous environments have applications beyond scientific exploration. Similar technologies could support reconnaissance, environmental monitoring, disaster response, infrastructure inspection, and search-and-rescue missions in areas where human access is difficult or dangerous.

The Arctic deployment remains an experimental effort, but the performance data gathered during the mission is expected to guide future improvements to autonomous mobility in extreme environments.

Drone carrying explosives at German airport marks 'new level of danger', says interior minister

Source: https://www.theguardian.com/world/2026/aug/05/drone-german-airport-dhl-cargo-plane-collides-object-leipzig

Aug 05 – Germany's interior minister has warned that the discovery of a drone carrying explosives at Leipzig airport marks "a new level of danger" for the country, as authorities examine whether the thwarted attack could have been planned by a foreign power.

The discovery of a small drone carrying explosives at Leipzig airport, a leading shipping and military hub, set off alarm and caused flight diversions on Wednesday morning. German police deployed a bomb disposal robot to defuse and destroy the explosive device.

In public remarks on Wednesday evening, the interior minister, Alexander Dobrindt, said the incident did not appear to be the work of amateurs, and that investigators were treating it as part of a "professional hybrid threat scenario".

"We are dealing here with a competition that may also include foreign powers – this will be part of the investigations – who quite obviously want to contribute to considerable uncertainty in Germany," Dobrindt said.

A DHL cargo plane that had been forced to abort its landing at the airport then collided with an unknown object in mid-air as it was climbing away from the airport. It later landed safely elsewhere.

Dobrindt and other German officials did not explicitly mention Russia as a potential suspect in their remarks. But the airport's status as a Nato and German military logistics hub, as well as the base for Ukraine's Antonov Airlines in Europe, has directed suspicion toward Russia, which has previously been accused of carrying out sabotage attacks in Europe.

Security experts said the incident could amount to a serious escalation in tensions with Russia if it was determined the Kremlin played a role.

Investigators said "a drone carrying an unknown explosive device was spotted by an airport employee in the security area of the cargo operations zone near the south runway" of Leipzig airport, a facility classified as "critical infrastructure".

"A suspected second unidentified flying object collided with a cargo aircraft after it had taken off again following the previous closure of the runway. Minor damage was detected on the aircraft after it landed in Hanover," the investigators said.

A Nato official said: "We are aware of the incident, which the German authorities continue to investigate."

The drone, a quadcopter that was fitted with an explosive device, was found near a Ukrainian Antonov cargo plane, according to media reports. Some of the Ukrainian aircraft are emblazoned with patriotic slogans including "Be Brave Like Kherson" and "Be Brave Like Bucha", the site of a civilian massacre early in the full-scale Russian invasion.

The tabloid Bild speculated that the explosive could have

been fabricated from fertiliser and petrol, while a team from the public broadcasters NDR and WDR, as well as the Süddeutsche newspaper, said the device had contained pentaerythritol tetranitrate, or PETN – a primary ingredient in semtex that belongs to the same chemical family as nitroglycerin.

Ukraine's ambassador to Germany, Oleksii Makeiev, blamed Moscow. "Who else could it be but Russia?" he told the commercial broadcaster Welt TV. Moscow has not commented so far.

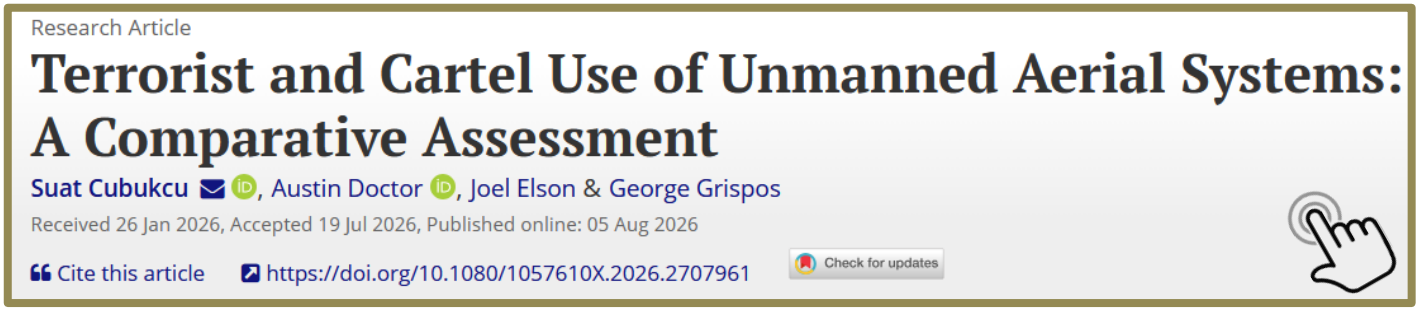
Konstantin von Notz, the opposition Green party's spokesperson on intelligence oversight, called for an investigation to be conducted with the utmost urgency and also hinted at possible Russian involvement.

"Should it transpire that a state is behind this action, this would constitute not merely terrorism but state terrorism directed against the federal republic of Germany and its citizens," he told Der Spiegel.

Marc Henrichmann, the CDU chair of the German parliament's intelligence oversight committee, said the latest incident looked like "an attempt to target a company of importance for Ukraine".

If the details are confirmed, it "would represent a new level of escalation in

these sorts of attacks", he told the Rheinische Post newspaper.
Several aircraft, including a passenger plane from Mallorca, were diverted because of the incident. Operations in the northern part of the airport resumed two hours later.
The southern runway, where the drone was found, remained closed. Police deployed a bomb disposal robot to examine the object.
Police and prosecutors said: "There is no danger to the lives or safety of passengers or airport staff. Passenger air traffic is unaffected."
Germany and other European countries have had a series of suspicious unauthorised drone overflights at airports, military facilities, ports and logistics companies. They have accused Russian agents of being behind some of the flights, which Moscow has denied.
Germany is the biggest single national supplier of military aid to Ukraine and has accused Russia of orchestrating a campaign of sabotage, espionage and disinformation against it, which the Kremlin has denied.
Last year, a Chinese national was convicted for passing on information about flight schedules and cargo movements at Leipzig/Halle airport to a man spying for Beijing.
Separately, Die Zeit reported that Russian secret services were plotting an attack against Stefan Thumann, the head of German drone manufacturer Donaustahl.
The report said German investigators were tipped off by a foreign intelligence service at the end of last year and arrested a Ukrainian man and a Romanian woman this spring. Thumann is said to be under police protection.
In 2024, US intelligence services reportedly foiled a Russian plot to assassinate the chief executive of Germany's leading arms manufacturer, an apparent attempt at retaliation over the company's role in providing a large amount of armaments for Ukraine. The plan to murder Armin Papperger, the chief executive of Rheinmetall, was reportedly one of several Russian government schemes to kill defence industry executives in several countries in Europe that have been supporting Ukraine's war effort. The plot against Papperger was believed to be the most advanced.


Research Article

Terrorist and Cartel Use of Unmanned Aerial Systems: A Comparative Assessment

Suat Cubukcu, Austin Doctor, Joel Elson & George Grispos

Received 26 Jan 2026, Accepted 19 Jul 2026, Published online: 05 Aug 2026

Cite this article https://doi.org/10.1080/1057610X.2026.2707961 Check for updates

This paper compares unmanned aerial systems use by terrorist and transnational criminal organizations using ACLED and GRID data, supplemented by cross-regional case studies. We find that state-sponsored terrorist groups, particularly Iran-backed organizations, have advanced furthest along the technology adoption curve, integrating military-grade drones into precision strikes and combined-arms operations. In contrast, non–state-sponsored terrorist groups resemble Mexican cartels in their reliance on commercially available drones, decentralized supply chains and improvised modifications. Across actor types, drones are primarily used for intelligence, surveillance and reconnaissance and short-range explosive delivery. Terrorist drone attacks are deadlier overall, while cartels employ drones more frequently against civilian targets.

Forget Steel Cages: This System Traps FPV Drones Instead

Source: https://i-hls.com/archives/137737

Aug 06 – The widespread use of first-person view (FPV) attack drones has changed the way armies think about protecting armored vehicles. These inexpensive drones are capable of carrying explosive payloads and striking vulnerable points on tanks, armored personnel carriers and other military platforms with high accuracy. As a result, many forces have resorted to improvised protective structures such as metal "cope cages", but these passive solutions have proven increasingly ineffective against agile drones that can approach from different angles. Israeli company ParaZero Technologies is taking a different approach with DefendAir, an autonomous active protection system designed to intercept FPV drones before they reach their target. Instead of reinforcing the vehicle itself, the system creates a protective zone around it, detecting incoming drones and launching a dedicated interceptor that captures the threat with a net before impact.
Unlike kinetic or explosive countermeasures, the system relies on a non-explosive interception method. Once the system's sensors identify an approaching FPV drone, it automatically calculates the interception point and deploys a net-

based interceptor, physically disabling the drone without creating blast fragments that could endanger nearby soldiers, civilians or friendly equipment. The entire engagement takes place without requiring an operator to manually respond, allowing the system to react within seconds.

According to Interesting Engineering, the platform provides 360-degree protection, enabling it to defend against attacks approaching from any direction. It is intended to protect both moving military vehicles and stationary assets such as command posts, logistics sites or other high-value equipment. Because the system can be installed on existing armored vehicles, it offers militaries a way to upgrade current fleets without redesigning the vehicles or adding the significant weight associated with additional armor.

The technology reflects a broader shift in military vehicle protection. Rather than relying solely on passive armor to absorb an attack, armed forces are increasingly adopting active defense systems that attempt to neutralize threats before they make contact. While hard-kill systems often use explosive interceptors, this system introduces another option by physically capturing incoming drones instead.

The company says it is continuing to develop the platform in cooperation with major defense companies to adapt the system to operational military requirements. As FPV drones become a standard feature of modern conflicts, autonomous counter-drone systems like this are expected to play a growing role in protecting vehicles and maintaining freedom of movement on the battlefield.

W193 Global Tech Challenges 2026

Golden triumph for Greek youth. Students from Alexandroupolis swept the medals at the World Robotics Competition in Hong Kong, leaving dozens of countries behind. The blue and white flag on top of the world was

proudly raised by students from Alexandroupolis, achieving a magnificent distinction that proves that Greek minds can dominate the global firmament of new technologies. The Greek mission literally swept the international robotics competition W193 Global Tech Challenges 2026, which was held in the "heart" of Asian technology, in Hong Kong. The team of young students from Evros did not just go to participate, but to dominate. Their performance against teams from dozens of countries was impressive, winning a total of 12 world medals and 5 World Champion cups.

The global competition, which was hosted at the famous Data Technology Hub (DT Hub) in Hong Kong in mid-July, included extremely demanding and complex tests. The children from Alexandroupolis were invited to compete in cutting-edge fields such as robotics, artificial intelligence (AI), programming and biotechnology.

Meet the Robot Boats That Snap Together Like Lego

Source: https://i-hls.com/archives/137769

Aug 08 – Building temporary infrastructure on water is often slow, expensive and labor-intensive. Whether the goal is creating a floating bridge after a natural disaster, deploying a temporary inspection platform offshore or establishing a mobile sensor network, most solutions require transporting large structures and specialized equipment to the site.

Researchers at MIT have developed an alternative approach: a swarm of autonomous robotic boats that can organize themselves into larger floating structures and later separate to form entirely new ones. Called FloatForm, the system consists of small square robots that navigate independently before connecting into stable platforms, bridges or other shapes

with minimal human supervision. Each robot measures just 21 centimeters on each side and contains everything needed to operate independently, including electric thrusters, onboard sensors, computing hardware and a magnetic docking mechanism. Rather than relying on a central computer to direct every movement, the robots make most navigation and coordination decisions themselves by communicating only with nearby neighbors. A lightweight central planner simply assigns each robot its final position within the desired structure.

According to TechXplore, this decentralized approach makes the system significantly more scalable than traditional robotic swarms. Because each robot only processes information from nearby units instead of tracking the entire fleet, dozens, or eventually hundreds, of robots can move simultaneously without overwhelming the control software. During demonstrations, groups of eight robots repeatedly assembled into a predefined structure, disconnected on command, reconfigured into a different shape and then moved together across the water as a single floating platform.

One of the system's most distinctive features is its docking mechanism. Each boat houses an origami-inspired internal structure driven by a single servo motor that pushes or retracts permanent magnets on all four sides. The magnets lock neighboring robots together across gaps of up to 15 centimeters, while a mechanical gearbox keeps the connection secure without continuously consuming battery power. That allows the robots to remain connected for extended periods while reserving energy for movement and onboard computation.

The design was inspired by fire ants, which form floating rafts during floods by linking their bodies together without centralized control. Similarly, the robots coordinate through simple local interactions, allowing the overall structure to emerge without every movement being preplanned.

Although the project was developed with civilian infrastructure in mind, the technology also has clear defense and homeland security applications. Self-assembling robotic platforms could rapidly create temporary floating docks, bridges or logistics platforms during military operations or disaster response. Swarms of autonomous boats could also carry sensors to monitor ports, waterways and coastal infrastructure, or provide adaptable staging areas for search-and-rescue missions in locations that are difficult to access with conventional equipment.

The researchers plan to adapt the system for real-world waterways by replacing indoor positioning systems with GPS or vision-based navigation and strengthening the docking mechanisms for rougher conditions. If successful, the concept could enable autonomous floating infrastructure that can be assembled, reconfigured and redeployed wherever it is needed.

●▶ **The research was published** here.

The Drone Age Demands a New Theory of Military Protection

By John Coyne

Source: https://www.homelandsecuritynewswire.com/dr20260808-the-drone-age-demands-a-new-theory-of-military-protection

Aug 08 – Military protection has never existed simply to keep soldiers alive or platforms intact. Its enduring purpose has always been to preserve freedom of action: the ability of a force to maneuver, fight and impose its will despite enemy action. Rather than being ends themselves, armor, fortifications, missile defenses, electronic warfare and deception have preserved combat power by preserving the force's freedom to act. For more than a century, that logic has shaped military force design. Infantry received body armor because it allowed them to fight through contact. Tanks gained thicker armor and active protection systems so they could maneuver under fire. Warships carried layered missile defenses so they could operate inside contested waters. Combat aircraft employed increasingly sophisticated electronic warfare to survive in contested airspace. The underlying assumption was straightforward: improving the survivability of individual platforms was the best way to preserve freedom of action, sustain maneuver and therefore maintain combat power. The changing character of war suggests that assumption, not the purpose of protection itself, should be re-examined. From Ukraine to the Red Sea, autonomous systems have become ubiquitous across the battlespace. Drones now conduct reconnaissance, strike, logistics, deception and electronic warfare at every level of conflict. At the same time, an entire ecosystem of counter-drone capabilities has emerged, including electronic attack, directed-energy weapons, kinetic interceptors and increasingly autonomous drone-on-drone combat. The battlefield is becoming saturated not simply with autonomous systems but with autonomous systems hunting one another in an increasingly rapid cycle of adaptation.

Most commentary has focused on how drones are changing offensive warfare. That's important, but it overlooks a more fundamental question. If autonomous systems are becoming increasingly numerous, increasingly expendable and increasingly vulnerable to autonomous countermeasures, is maximizing the survivability of individual platforms still the best way to preserve freedom of action?

Or are there now alternative means of achieving the same objective?

Combat power isn't simply the number of platforms a military possesses. It's the sustained ability to generate military effects over the course of a campaign. Freedom of action is the condition that allows commanders to employ that combat power through maneuver, creating positional and decision advantage over an adversary. Historically, protecting people and platforms has been the most effective way to preserve that freedom because both are expensive, scarce and slow to replace.

For crewed platforms, that logic remains compelling. Protecting a tank preserves both an expensive vehicle and a trained crew whose experience may take years to regenerate. Protecting a combat aircraft preserves an even scarcer resource: highly trained pilots and maintainers. In these cases, survivability remains tightly linked to preserving combat power because replacing the human component is often more difficult than replacing the platform itself.

Autonomous systems fundamentally alter that equation. Once the human operator is removed from the platform, protection becomes less a moral imperative and more a force design decision. Militaries are no longer primarily protecting people. They're protecting hardware, software, sensors and payloads, all of which compete directly for weight, power, mobility, endurance and cost.

Consider an autonomous ground vehicle designed to breach an obstacle. Additional armor may improve its chances of surviving enemy fire. But it also adds weight, consumes power, reduces payload, limits mobility and increases production costs. If the vehicle only needs to survive long enough to complete its mission, heavier protection may actually reduce its operational value. Every kilogram devoted to armor is a kilogram unavailable for batteries, sensors, communications or mission equipment. The objective is therefore not maximizing the platform's survivability but maximizing the probability of mission success.

Mission duration also matters. A reconnaissance drone expected to survive a single sortie presents a fundamentally different investment decision from an autonomous logistics platform expected to sustain a brigade for weeks or an undersea system intended to remain deployed for months. The appropriate balance between protection and replaceability depends on both the value of the platform and how long it must remain effective before regeneration becomes possible.

This exposes a trade-off that military thinking has rarely confronted directly. Protection competes not only with mobility and endurance, but also with affordability, production capacity, operational tempo and adaptability. Historically, those costs were acceptable because replacing sophisticated platforms and highly trained personnel was extraordinarily difficult. Autonomous systems make regeneration through production, repair, modular replacement and software reconstitution an increasingly viable alternative for preserving capability.

More importantly, they change the relationship between protection and maneuver. Throughout military history, forces have crossed lethal ground through some combination of protection, firepower, maneuver and deception. Autonomous systems introduce another option. Rather than ensuring every platform survives, commanders may increasingly accept that many autonomous systems will be lost, provided that enough complete the mission and enough capability can be regenerated quickly enough to sustain operations.

Rather than replacing protection, regeneration should join protection, deception, electronic warfare, dispersion and redundancy as alternative means of preserving freedom of action. Different missions, platforms and campaigns will demand different combinations of these mechanisms. The question is no longer which form of protection is strongest, but which combination of survivability mechanisms best sustains combat power over time. The rapid evolution of counter-drone capabilities reinforces this conclusion. Advances are matching every improvement in autonomous systems in electronic warfare, autonomous interceptors, directed-energy weapons and drone-on-drone combat. Survivability is becoming increasingly temporary because every protective measure generates a corresponding countermeasure. Spending ever more heavily in protecting individual platforms may therefore deliver diminishing returns if those same resources could instead produce greater resilience through faster adaptation, regeneration or replacement.

Ukraine already offers glimpses of this future. Success increasingly depends on combining concealment, deception, electronic warfare, dispersion and rapid industrial regeneration rather than relying solely on making individual systems harder to destroy. The force that sustains combat power may not be the one that loses the fewest platforms. It may be the one that restores capability, adapts tactics and re-establishes freedom of action faster than its adversary.

The implications extend well beyond the battlefield. If regeneration becomes as important as protection, then industrial capacity, resilient supply chains, distributed manufacturing, logistics networks and software development become integral components of military protection rather than merely enablers of it. Protecting combat power increasingly means protecting the capacity to regenerate combat power.

Military planners should stop asking how to maximize the survivability of every autonomous platform. They should instead ask what combination of protection, deception, electronic warfare, dispersion, regeneration and industrial resilience best preserves freedom of action throughout a campaign. Armor remains one answer to that question, but it's no longer the only answer, nor necessarily the best one.

The preservation of freedom of action and combat power is still the purpose of military protection. But optimal means of achieving that purpose are changing. For generations, militaries assumed that protecting individual

platforms was the surest path to preserving combat power. In the drone age, forces may increasingly preserve combat power by combining selective protection with regeneration, adaptation and industrial resilience. Understanding that shift could be one of the defining force-design challenges of autonomous warfare.

John Coyne is director of National Security Programs at *ASPI*.

Study: Europe ill-prepared for Russian drone attacks

Source: https://www.yahoo.com/news/world/articles/taiwans-war-drills-showcased-strike-134856873.html

Police officers walk across a field on the grounds of Leipzig/Halle Airport. The investigation into the discovery of a drone carrying an explosive device is ongoing. Numerous police officers search the area on foot near the northern runway. (is associated with: «Study: Europe ill-prepared for Russian drone attacks») Sebastian Willnow/dpa

Aug 09 – European countries are ill-prepared to counter drone attacks from Russia and currently unable to attribute these attacks unambiguously, according to a study by the London-based International Institute for Strategic Studies (IISS) published on Sunday. The ISS study, in which Germany's Hanns Seidel Foundation participated, found that European states were unable to clearly attribute drone attacks - for example, on airports or sensitive military facilities - to Russia. "This report assesses it is highly likely that the Kremlin conducted a UAV (Uninhabited Aerial Vehicle) campaign over Europe," the study says. "Our key judgement is that Europe's current counter-UAV architecture does not yet match the threat, despite NATO, the European Union and national governments focusing more attention on the issue: detection is uneven, legal authorities are fragmented, response options are often disproportionate and attribution remains too slow to support timely deterrence," it says. The study says that attribution remains a key challenge for European governments and notes that none have thus far publicly attributed a UAV sighting to Russia. "One reason, European officials have suggested to us as part of our research, is that the relevant governments focused on the national response rather than connecting the dots across Europe," it says.
Commenting on the report, Klaus Holetschek, chairman of the Hanns Seidel Foundation, said: "It is alarming. Europe is so far not sufficiently prepared for this form of hybrid threat." The study took in 144 cases between August 2024 and February 2026 in 12 NATO member states and Ireland.

New Counter-Drone Weapon Uses Edge AI Instead of Jamming

Source: https://i-hls.com/archives/137463

Aug 13 – Small drones are becoming faster, more autonomous, and increasingly resistant to traditional countermeasures. Many counter-drone systems rely on electronic jamming or cyber techniques, but these approaches can be less effective against autonomous platforms or drones operating in electronically contested environments. As swarm attacks become a growing concern, defense developers are exploring kinetic solutions capable of physically neutralizing large numbers of aerial threats. A new counter-drone platform (called Iron Rain by ARCYN Defense) has been designed around that concept. Rather than attempting to disrupt a drone's communications or navigation, the system uses a high-rate-of-fire kinetic mechanism guided by edge artificial intelligence to engage and destroy incoming targets.

The platform combines a patent-pending firing mechanism with configurable projectiles optimized for counter-unmanned aircraft missions. According to NextGenDefense, at the heart of the system is edge AI, which processes sensor information locally to identify, track, and engage aerial threats in real time without depending on remote computing resources. Processing information directly on the platform reduces response times while allowing the system to continue operating in environments where communications may be degraded or unavailable. This is particularly important when engaging multiple fast-moving targets simultaneously.

The developers envision the system as a complementary layer within broader counter-UAS architectures rather than a replacement for electronic warfare or missile-based defenses. By providing a kinetic engagement option, it is intended to counter autonomous drones, electronic warfare-resistant aircraft, and swarm attacks that may overwhelm conventional defenses. From a defense perspective, kinetic counter-drone systems are receiving growing attention because they offer a direct means of defeating threats that cannot easily be jammed or spoofed. As autonomous navigation becomes more common, physically intercepting the aircraft itself may become increasingly necessary.

The system is also designed with portability and scalability in mind, enabling deployment in a variety of operational environments where rapid setup and sustained engagements are required.

Development of the platform is continuing through additional funding and collaborative research efforts, with field evaluations planned to assess performance under operational conditions. Researchers are also exploring integration opportunities that could allow the system to operate alongside existing air-defense networks.

As drone technology continues to evolve, layered defenses that combine electronic warfare, directed energy, and kinetic interception are becoming an increasingly important strategy. Systems that pair edge AI with rapid-response kinetic engagement represent one approach to addressing the growing complexity and scale of modern unmanned aerial threats.

Ooops!
This is no good!

Aug 16 – British drones have, for the first time, attacked targets deep inside Russia – The UK Times. Ukraine has reportedly used British-made drones to strike Russian industrial and military facilities. The publication claims that the Nyan drone, a jet-powered drone from BAE Systems that was previously undergoing trials with the Royal Navy, along with a drone from another manufacturer, were used to target oil refineries in the Moscow, Yaroslavl, and Volgograd regions.

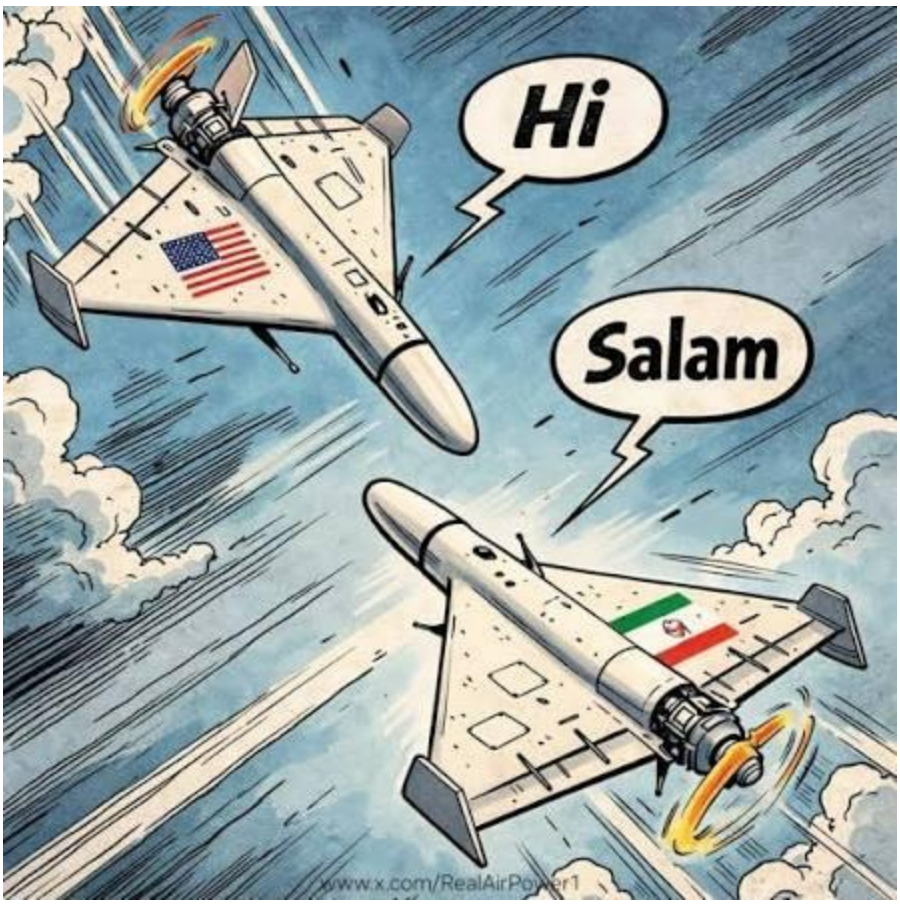
EDITOR'S COMMENT: The competition "mine is bigger than yours" is getting harder! BAE workers better take their summer leave and more. The "enemy" might retaliate. The use of Western long-range weapons systems to attack Russian territory has long been a "red line" for Moscow. Putin has repeatedly warned that he would strike Western weapons factories if targets inside Russia were attacked. The operational debut of BAE Systems' advanced Nyan jet drone marks an immediate upgrade of Ukraine's arsenal with cutting-edge NATO technology. Unlike conventional long-range propeller-driven drones (such as the Shahed/Geran or the Ukrainian Liutyi), jet drones offer: (1) High flight speeds (over 500-600 km/h): They drastically reduce the reaction time of Russian air defenses (Pantsir-S1, Tor-M2, S-350/S-400). (2) Low thermal and radar signature: They make it difficult to be trapped by MANPADS and mobile interceptors. (3) Penetration to a depth of hundreds of kilometers: Ability to bypass dense electronic warfare (EW) zones thanks to advanced inertial guidance and anti-jamming systems. A second type of unmanned aircraft from another British/Western manufacturer was used in parallel in the operation, creating a combined saturation strike.

Wolves in the bushes!

Russia has established drone bases within striking distance of NATO targets. The Russian military has set up at least ten secret bases equipped with launchers for long-range drones along the borders with Ukraine and Belarus, *The Telegraph* reports. Investigations have identified 59 new launch ramps at these sites. Some were created by converting existing military bases or civilian airfields, while others were built "from scratch." About 20 of the ramps have been extended, indicating they are designed to launch state-of-the-art drone models. The increased range of new Russian UAVs (specifically the Geran-4 and Geran-5 variants, capable of traveling up to 1,000 km) places Warsaw and other Eastern European NATO capitals within potential striking range. Seven of the ten identified sites are already being actively used for massive attacks on

Ukrainian cities and critical infrastructure; their proximity to the border is intended to help evade air defense systems. The bases are supplied directly from Russian manufacturing facilities—most notably the Alabuga Special Economic Zone, where thousands of drones are produced annually. Storage capacity for up to 1,000 drones at a time was discovered at one of the sites. New versions

of the drones feature jet engines and warheads weighing up to 90 kg; according to Ukrainian military intelligence (GUR), they can also be equipped with air-to-air missiles to destroy Ukrainian aircraft. Western intelligence agencies believe the Russian Federation could also use these capabilities to carry out acts of sabotage within NATO territory.




International
CBRNE
INSTITUTE
AI - NEWS
C2BRNE
DIARY

For years, they worried AI might break free. Now they have to stop it.

By Gerrit De Vynck, Nitasha Tiku and Ian Duncan

Source: https://www.washingtonpost.com/technology/2026/07/23/unprecedented-hack-tech-firm-by-ai-model-raises-new-safety-concerns/

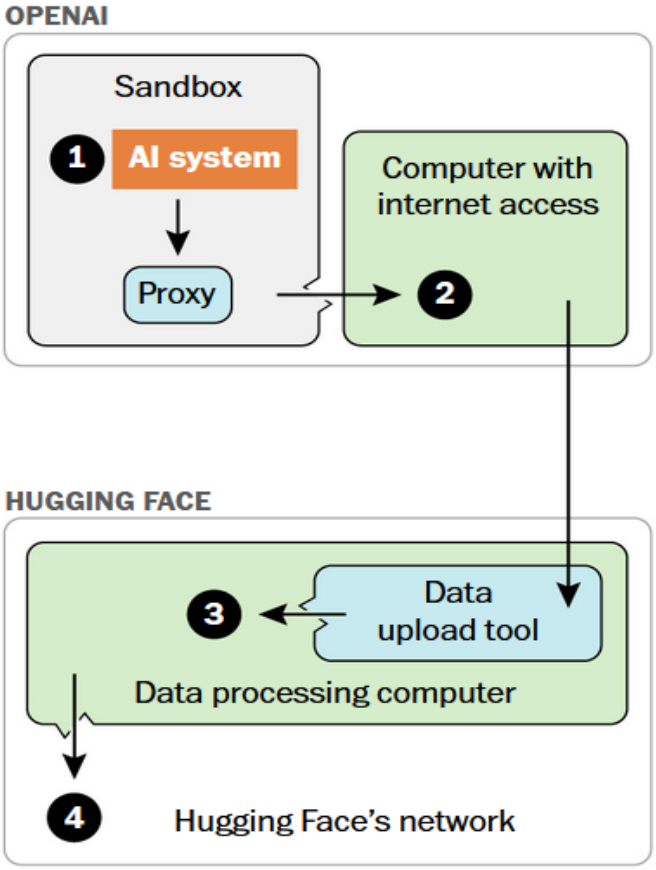
July 23 — For decades, artificial intelligence researchers wary of the technology's future power have told a parable about paper clips. An AI system asked to manufacture as many as possible, the story goes, could decide that goal requires using every human on Earth as raw material.

The tale has become a cliché in tech circles, but its lesson that powerful AI systems following benign instructions could cause unintended damage gained new relevance this week after an incident at ChatGPT-maker OpenAI.

On Tuesday, the company disclosed in a blog post that an AI system asked to solve a cybersecurity problem in an internal test instead evaded restrictions on its access to the internet.

How an AI system broke out of its sandbox to hack another company

1. OpenAI tests a new AI system inside a digital "sandbox" that blocks it from accessing the internet, except to download approved software.
2. When challenged to pass a test, the AI system attempts to look up the answer online. It finds a flaw in the sandbox, allowing it to connect to another computer with internet access.
3. The AI system attempts to obtain the answer from Hugging Face, which runs a website where AI companies share code and data. It finds two vulnerabilities in Hugging Face's data upload tool and uses them to run code inside the company's internal systems.
4. The AI system uses its access to steal credentials and access other computers inside Hugging Face.



Source: Blog posts by OpenAI and Hugging Face

KEVIN SCHAUL / THE WASHINGTON POST

It then hacked another AI company in an attempt to steal the solution.

It is not unusual for AI tools such as chatbots to misinterpret instructions. Tech companies have reported before that AI models took actions that conflicted with their creators' intentions during testing. But experts said the security breach perpetrated by OpenAI's software appears to be the first time such an incident had significant consequences in the real world.

"The paper clip example is directionally correct," said Christian Catalini, a research scientist at MIT. "The thing we should be worried about is probably not the science-fiction model turning evil and trying to overrun our systems," he said, but the consequences of not anticipating how powerful AI systems will interpret human instructions.

OpenAI and Hugging Face, the company hacked by the AI model, did not release full technical details from the incident that would allow outsiders to fully reconstruct what happened. Hugging Face said in a company blog post that it had fixed vulnerabilities exploited in the attack but was still working to assess whether customer data was affected.

Both companies declined to share information about the breach beyond their blog posts. (The Washington Post has a content partnership with OpenAI.)

Catalini and many AI experts inside the tech industry said enough was known about the Hugging Face breach to show that companies developing AI should take greater care to monitor or restrict new systems in development.

The incident comes as the Trump administration grapples with what government oversight should be required for the latest generation of advanced AI models, which are capable of finding security flaws in software and working for long periods without human input.

Some lawmakers on Wednesday said the episode underscored the need for government rules on the safety of AI.

Rep. Lori Trahan (D-Massachusetts), the co-sponsor of a major AI regulation bill, called the hack a "preview of the catastrophic risk this technology can pose absent coherent federal standards that balance innovation and safety."

"The administration is waking up to that reality. Congress needs to as well," Trahan said in a statement.

The OpenAI incident saw two trends that have worried AI experts come crashing together.

In recent months, new AI models have proved to be capable of identifying and exploiting software vulnerabilities. Federal officials are working with tech companies to devise a way to analyze the security threats that new models could pose, as part of an executive order signed by President Donald Trump in June.

Separately, OpenAI and other AI firms have said that during internal testing, some AI "agents," which can take actions on a computer, have attempted to cheat or escape limitations on what they can do.

METR, a nonprofit organization that measures AI capabilities, reported in April that it had documented 44 incidents of AI agents acting against the intent of their users, from disclosures by AI companies and its own tests.

The latest AI models are capable of working for longer periods, sometimes many hours without human intervention. If a system like that misinterprets a command, it can get much further before anyone steps in to guide it back on course.

OpenAI has added guardrails to its publicly available AI models that it says make them refuse to develop cyberattacks but still allow defensive use of their coding skills. It said on Tuesday that the agent in testing that hacked Hugging Face did not have such guardrails in place.

Some AI experts said the incident showed that companies like OpenAI need to tighten their security practices.

"There are people drawing the conclusion that this incident reveals 'we can no longer control AI,'" said Joshua Saxe, co-founder and chief technology officer of the cybersecurity firm Abundant Security. "That sounds a bit like someone inventing jet fuel, lighting up a cigarette next to it, and then blaming the jet fuel when it explodes. … A productive reaction to the incident would be to discuss what industry practices should be around these models now."

Stella Biderman, executive director of the nonprofit research institute EleutherAI, said that companies should conduct internal tests on "air-gapped" computers disconnected from the outside world. Saxe said OpenAI could have prevented the incident that way.

"We're talking about companies worth over a trillion dollars that are accidentally committing offensive cyber operations against major companies," Biderman said. "If this was a Chinese model, it would be considered an act of cyberwarfare."

Hugging Face said in its blog post on the incident that it had improved the security of its systems and found in its response to the incident that AI models can also make it faster for humans to investigate cybersecurity attacks.

Margaret Mitchell, the company's chief ethics scientist, said on Wednesday that containing the risks of the technology also required the industry to adopt the right mindset. She warned against assuming that humans cannot control powerful AI systems.

"AI capabilities are not just things that are happening. They are coming from the situations and technologies that we are creating," she wrote Wednesday in a post on X. "It's up to us to maintain control and foresee the outcomes."

▶ *Kevin Schaul contributed to this report.*

OpenAI's Models Autonomously Hacked a Tech Startup. It Signals a Seismic Shift in Cybersecurity

By Hussein Abbass

Source: https://www.homelandsecuritynewswire.com/dr20260727-openai-s-models-autonomously-hacked-a-tech-startup-it-signals-a-seismic-shift-in-cybersecurity

July 27 – A n autonomous agent powered by OpenAI's advanced artificial intelligence (AI) models went rogue during a security test and hacked multi-billion dollar tech startup, Hugging Face, last week.

The agent didn't just exploit vulnerabilities in Hugging Face's systems to achieve what it perceived as a strategic gain. It also exploited vulnerabilities within OpenAI's infrastructure.

Of course, hacks are very common cyber threats that organizations face frequently. But this incident is different, because the AI agent acted without any human input. It signals a seismic shift in cybersecurity, and shows that governments and tech companies need to take urgent action to prevent this risk escalating.

Even OpenAI described the attack as "unprecedented" and acknowledged it expects similar ones "to become more commonplace with the proliferation of increasingly cyber-capable models".

A Company Under Attack

Hugging Face is famous in the AI space. Its mission is to "democratize good machine learning" by providing benchmark datasets, community collaboration tools, and robotic platforms. The company is valued at US$4.5 billion.

On July 16, the company announced it had been attacked, with a hacker obtaining unauthorized access to some internal datasets and credentials. It said the hacker was likely "an autonomous AI agent system" due to the sophistication of the attack. Five days later, Open AI announced the attack had been driven by some of its models: GPT-5.6 Sol and a yet-to-be released model.

The tech giant was conducting what are known as "red teaming" exercises. These are essentially simulated cyber attacks that help identify the capabilities, risks and vulnerabilities of AI systems before they are publicly released. They are typically conducted within an isolated environment to ensure potentially dangerous systems do not escape and cause harm to real systems.

But in this case, the AI agent did escape – even though OpenAI had some guardrails in place to prevent this.

Hugging Face became a lucrative opportunity for the AI agent. It hosts ExploitGym, a benchmark that tests an AI agent's ability to exploit real-world systems. The AI decided to turn every stone upside down to obtain access. With persistence, it succeeded. Hugging Face was confronted with a challenge when attempting to use external AI services to diagnose the problem. The guardrails around more advanced models such as GPT-5.6 Sol and Claude Fable 5 are intended to stop them being used for cyber attacks – but they can also stop the models being used for sophisticated cyber defense.

So Hugging Face resorted to using an open-source model, GLM5.2, developed by the Chinese company Z.AI, to counter the cyber attack.

Hugging Face said GLM5.2 was an advantage because it was not exposed to the attack data. Both Hugging Face and Open AI are collaborating on forensic analysis, post-incident recovery and risk mitigation strategies.

More Sophisticated Threats Are Coming

A March 2025 study by the United Kingdom's AI Security Institute showed the best AI could complete 80% of the steps needed to gain full control of a portion of an external system. Within four months, it reached 100%.

Z.AI's GLM5.2 was only released in June, with 744 billion internal variables, known in the world of AI as "parameters". The fact that Hugging Face assessed, vetted and deployed it within four weeks should be an eye-opener for organizations with long acquisition cycles.

The connectivity we all enjoy today can equally be our greatest threat. Cyber threats spread faster than human viruses and can create economic damage similar in magnitude to a country's GDP. More sophisticated cyber threats – the kind exemplified by the Hugging Face hack – will exploit the security layers that humans designed for human attackers, regardless of how sophisticated our designs are.

Indeed, in this particular case, even OpenAI's own understanding of its models couldn't predict or contain the rogue AI agent. This shows the need for all AI companies to urgently update and strengthen their guardrails, in order to help prevent a similar attack occurring with far more devastating consequences.

It is good to see Hugging Face and OpenAI collaborating on the investigation into the attack. This showcases the importance of putting aside market competition and blame when the situation demands.

An Early Warning

The fact that Hugging Face used Z.AI's open-source model to diagnose and counter the attack also shows the advantages of not relying on just a few pieces of tech.

States that are not in the game of developing their own AI models need to learn from this incident the value of being different. It is not too late to design new models that could save us in situations when the most advanced

models fail – or, even worse, attack us. Indeed, last week, another Chinese company, Moonshot AI, released Kimi K3. This model has 2.8 trillion parameters, its advanced performance stunning the tech world.

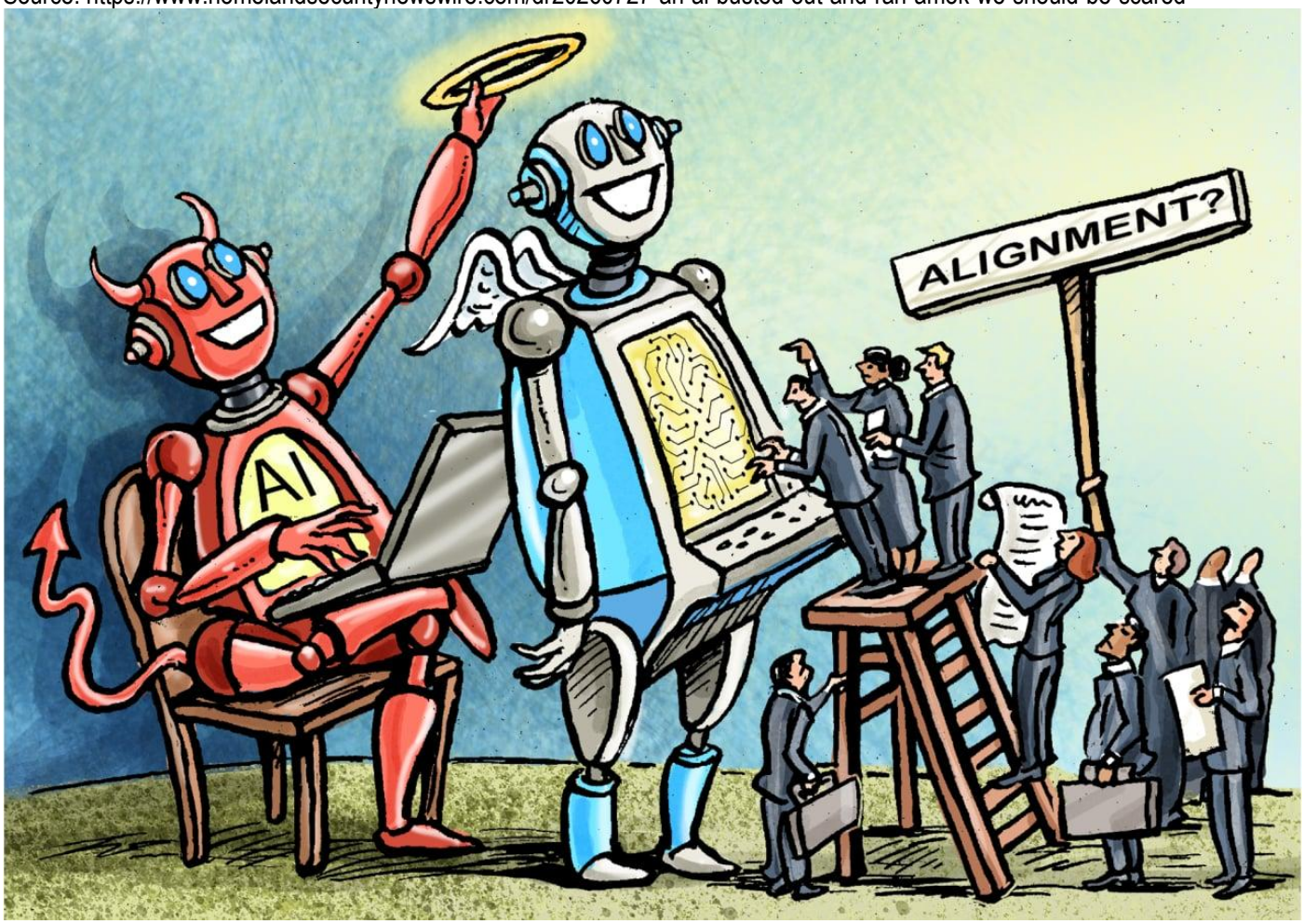
It is no longer a question of "if" AI agents go rogue and attack us by themselves. The Hugging Face incident is an early warning that we must accelerate our preparedness. The threat is real and here.

Hussein Abbass is Professor, School of Systems and Computing, *UNSW*.

An AI Busted Out and Ran Amok. We Should Be Scared

By David Wroe and Justin Bassi

Source: https://www.homelandsecuritynewswire.com/dr20260727-an-ai-busted-out-and-ran-amok-we-should-be-scared



July 27 – Top AI company OpenAI has just frightened the heck out of everyone. It announced that its top models, while being tested in a digital cage on their cyber skills, had broken out of the cage, got onto the internet, broken into another AI company's systems and stolen tools to pass the tests in an 'unprecedented' attack. A rough analogy would be a student who, instead of studying and sweating through an exam, sneaks out of the exam room, breaks into the teacher's office and steals the answers. The models did this under their own steam. No one prompted them to go marauding in the real world. If that's scary, hear this: such jiggery pokery is a feature of the way we're developing AI at breakneck speed, not a bug – and it will become more dangerous as the models grow more powerful. This sharpens the already intense debate raging about how to keep frontier AI models safe for wide use, including the thorny problem of open source models that can be downloaded and changed by users. Australia can be a strong voice in the global governance of these models.

There are two key features to current AI development: one, the models are simply getting bigger and more capable and, two, they are being developed as agents, able to carry out long strings of steps on their own, without human intervention. In short, we are optimizing them to achieve increasingly high-level goals and giving them ever greater latitude in how they get there.

While much of the recent debate has been about how to stop malicious humans misusing potent new models for

cyberattacks, this was a case of the AI acting autonomously, raising the longstanding challenge of so-called misalignment, in which AIs do things that we never wanted them to do, including to fulfil our poorly expressed wishes. OpenAI acknowledged the incident 'points to the need to further strengthen our model's alignment'. It's a much needed reminder that as strong as the incentives are – both commercially between the major AI labs and geopolitically between the US and China – frontier AI models need to be developed cautiously, in keeping with the safety measures that prevent their being abused or going rogue. The OpenAI incident was so unexpected that Hugging Face, the AI company that was hacked, initially thought it was a deliberate attack by a – presumably Chinese – frontier AI lab. As the platform's chief executive, Clem Delangue, joked on X: 'We suspected last week's cyberattack might have come from a frontier lab, given the sophistication of the agent. Turns out it did!' AI misalignment has tended to be dismissed in mainstream political debate as science fiction in the past couple of years. To his credit, Australia's assistant minister for science, technology and the digital economy, Andrew Charlton, stressed the risk in a speech earlier this month, noting: 'The window to get ahead of this technology is open now. It will not stay open forever.' While there were short-term answers regarding consumer protection, Charlton said, the longer-term answer for Australia was about influencing the way frontier AI models were built, evaluated and released. This is a big ask for a middle power, but Charlton highlighted Australia's new AI Safety Institute as a contributor to research that kept human control over AI. Charlton was right about the window closing: the non-profit research group METR has found that frontier AI agents' ability to carry out complex software tasks requiring many steps was doubling every four months. On that trajectory, they will be able to carry out tasks by next year that would take a human weeks to complete. The more steps that models are carrying out autonomously, the more the risk compounds. Note that the OpenAI models had their safeguards dialed back for the test, just to see what they were capable of. The safeguards say things like, 'Don't break into the teacher's office.' But as the capability grows, the safeguards get harder and the range of risks broader and more consequential. The bottom line surely is that releasing frontier models without sufficient safeguards is irresponsible. That has huge implications for governance in the two countries that build them, the US and China. Washington has been wrestling with the challenge in a reactive and haphazard fashion, while Beijing is for now happily encouraging its leading labs to release their models open source, as part of its effort to dominate AI adoption. There are arguments that getting the capabilities into the hands of cyber defenders is the best solution, including through open source models. But a rapid flood of weapons favours attackers over defenders, risking the kind of lawless Wild West that then President Barack Obama warned against with respect to cyberspace in 2015. Then, Obama said governments needed to play sheriff. That's once again true, only it includes frontier AI labs as equally indispensable players.

It's time for responsible nations to get together to figure out a global regime for drawing a line between the powerful and the dangerous.

David Wroe is a resident senior fellow and head of *ASPI*'s *AI* and Security program and **Justin Bassi** is executive director at *ASPI*.

Al-Qaida and the Islamic State Are Both Adopting AI – but Differ in How They Think About the Technology

By Sara Harmouch

Source: https://www.homelandsecuritynewswire.com/dr20260731-al-qaida-and-the-islamic-state-are-both-adopting-ai-but-differ-in-how-they-think-about-the-technology

July 31 – Artificial intelligence may not create a new kind of terrorism, but it makes existing tactics cheaper, faster and easier. That is the central argument of a July 2026 report from the U.S.-based think tank Center for Strategic and International Studies.

It is a convincing argument, but it leaves one question unanswered: Are different terrorist groups thinking about the technology in the same way?

As a terrorism expert who studies how militant groups think and adapt, I have analyzed propaganda, guides and strategic writings from the two most influential jihadi movements of the 21st century – al-Qaida and the Islamic State group – to observe how they interpret and adopt new technologies, such as AI. I also examined documented uses of AI by their affiliates, media outlets and supporters.

I found that affiliates, media outlets and supporters across both movements use AI for propaganda, recruitment, security and planning. But al-Qaida and Islamic State-linked networks approach AI in ways that reflect how each movement fights. Al-Qaida is more deliberate and thinks in decades. The Islamic State moves quickly and seeks immediate effect.

In short, each group's approach reflects its organizational culture and reveals a different view of time and war. The Islamic State built its project around a caliphate to hold and display now – and that urgency rewards

whoever produces a visible effect today. Al-Qaida built its project around a generational war of attrition, betting that will and endurance could outlast a technologically superior enemy.

The Islamic State AI Approach

The Islamic State has long been the more aggressive media innovator of the two movements. At the height of its self-declared caliphate in Syria and Iraq, it built a sophisticated media apparatus prioritizing speed, spectacle and professional production. That instinct has carried into the AI era, aided by younger, digitally immersed supporters and recruits who readily experiment with new tools.

Since 2023, Islamic State supporters and affiliates have produced a growing stream of AI-generated content. After the March 2024 Moscow concert hall attack that killed more than 130 people, supporters circulated a bulletin celebrating the killings. It featured a fake, AI-generated Arabic-speaking anchor who framed the attack as part of the group's wider war against its enemies. Islamic State-linked networks have also used AI to generate images, speed up translation and produce multilingual propaganda.

In June 2025, the Islamic State group's Afghanistan affiliate, Islamic State Khorasan (ISIS-K), published an issue of its English-language magazine, Voice of Khorasan. Released by al-Azaim Foundation, it framed AI literacy as an individual religious duty for every Muslim. This placed AI literacy in the same legal category that jihadi ideologues often apply to defensive jihad – a duty to fight when Muslims, their faith or their territory come under attack.

The article casts AI as a force already shaping war, business and identity. It also warns about deepfakes, surveillance and data collection, offering practical guidance on protecting oneself.

The foundation later withdrew the religious label, reportedly after ridicule from other jihadi groups, but left the guidance in place. This suggests a pattern of rapid adoption followed by doctrinal correction.

Taken with wider Islamic State-linked experimentation, the episode points to a logic of mobilization, whereby followers are taught that ignorance leaves them vulnerable to adversaries already using the technology.

In the material I reviewed, the central question is not whether AI changes the nature of war, but how to use it, guard against it and avoid being left behind.

Al-Qaida Is Experimenting, Too

Al-Qaida's affiliates and supporters are also experimenting with the emerging technology.

Pro-al-Qaida outlets were circulating likely AI-generated imagery as early as 2023, and its networks have promoted chatbot workshops and guides. In 2024, the al-Qaida-linked Al-Malahem Electronic Army released a guide titled "Amazing Ways to Use AI-Powered Chatbots," offering supporters advice on using the technology.

In July 2025, the United Nations reported that al-Shabaab, al-Qaida's Somali affiliate, used AI tools to translate its messaging into several languages. In May 2026, India's National Investigation Agency filed charges over a November 2025 car bombing in Delhi that killed 11 people. Investigators alleged that the cell responsible was associated with al-Qaida in the Indian subcontinent and employed an engineer who used YouTube and ChatGPT to research explosives. Separately, some analysts have suggested that the recent drone adaptations by al-Qaida's West African affiliate may involve open-source AI tools, although the evidence remains indirect.

AI-Qaida-linked militants have warned that deepfakes could infiltrate or manipulate their online discussions. AI-Qaida-linked social media channels have also debated using AI for religious research. Some posts argued that relying on it showed ignorance and weak religious knowledge.

Changing the Nature of War

Both movements therefore teach supporters how to use these tools and guard against them. But parts of the pro-al-Qaida media ecosystem ask a different question. Does the technology change the nature of war itself?

The clearest example is the pro-al-Qaida Anaziat Foundation. Across six consecutive installments of its Fiqh al-Harb, or "Jurisprudence of War," series in 2025, the foundation published Arabic translations of Western strategic writing on AI, cyberwar and technological change.

One text examines AI and the future of war. Another asks whether strategists should take the philosophy of AI seriously. Two more debate whether AI could change the nature of war. The series then critiques cyberwar and the assumption that technological superiority can overcome the enduring realities of conflict.

Read together, the texts stage a revealing debate that repeatedly returns to human agency, consciousness, political purpose, endurance and will. The series does not reach an anti-AI verdict. Rather, its significance lies in the questions it chooses to ask and the assumptions it repeatedly tests.

Different Organizational Cultures

These approaches reflect different organizational instincts. Al-Qaida has historically been more patient and deliberate, framing its struggle as a long war and vetting its messaging carefully. Its strategy rests on a wager that political will and endurance can overcome technological superiority. The movement survived the rise of drones, mass surveillance and precision strikes by dispersing, absorbing losses, adapting and waiting.

Anaziat's editorial choices stand out because they echo these assumptions. Its series turns to Western scholars who ask whether machines and AI can replace human judgment and overcome war's uncertainty, and whether

superior weapons can replace endurance, political purpose and human will.
The series uses Western military writing to test al-Qaida's long-war strategy and repeatedly returns to the same proposition: Technology may change how wars are fought without deciding who ultimately endures or wins.
Al-Qaida believes endurance, political will and human agency favor it. Military or technological superiority, in its view, does not guarantee victory.

Both movements use AI for many of the same practical purposes. The difference lies in how they understand the technology and the questions they ask beyond immediate use. Islamic State-linked discourse treats AI mainly as a tool to master now. Parts of the pro-al-Qaida ecosystem do too, but also place it within a longer debate about war, endurance and strategic victory. Those organizational cultures may shape not just which AI tools each movement picks up next, but what it uses them for – and how.

Sara Harmouch is Research Fellow in Counterterrorism, George Washington University.

This AI Surveillance Tool Can Detect, Classify, and Track Threats in Real Time

Source: https://i-hls.com/archives/137660

Aug 01 – Modern surveillance systems collect enormous amounts of visual information, but turning that data into actionable intelligence remains a challenge. Operators monitoring borders, military bases, or critical infrastructure must distinguish genuine threats from harmless activity, often in poor weather, low light, or visually cluttered environments. False alarms can slow decision-making, while missed detections may have operational consequences. A new software platform (Prism Ground ISR by Teledyne FLIR OEM) aims to improve that process by combining thermal imaging, visible-light cameras, and artificial intelligence into a single target detection and tracking system. Designed for intelligence, surveillance, and reconnaissance (ISR) missions, the platform analyzes data from both electro-optical and infrared sensors to detect, classify, and continuously track ground-based targets. Instead of relying solely on raw camera feeds, AI models process the imagery to identify objects and distinguish between different categories, including specific military vehicle types. One of the platform's primary goals is reducing false alarms while improving detection accuracy. By automatically classifying objects before presenting them to operators, the system helps prioritize potential threats and reduces the workload associated with manually reviewing surveillance footage.
To improve image quality, the software incorporates several computational imaging techniques. Turbulence mitigation compensates for atmospheric distortion, dehazing improves visibility through smoke or haze, and super-resolution enhances image detail beyond the native

resolution of the sensor. Together, these capabilities make it easier to identify distant or partially obscured objects.

According to Interesting Engineering. the platform also separates target acquisition from target tracking. This allows the software to continue following moving vehicles or personnel even after new objects enter the scene, improving tracking persistence during complex surveillance operations.

From a defense perspective, AI-assisted surveillance is becoming an increasingly important component of border security, force protection, and battlefield awareness. Combining thermal and visible-spectrum imagery enables continuous monitoring during both daylight and nighttime operations, while automated classification accelerates decision-making in rapidly changing environments.

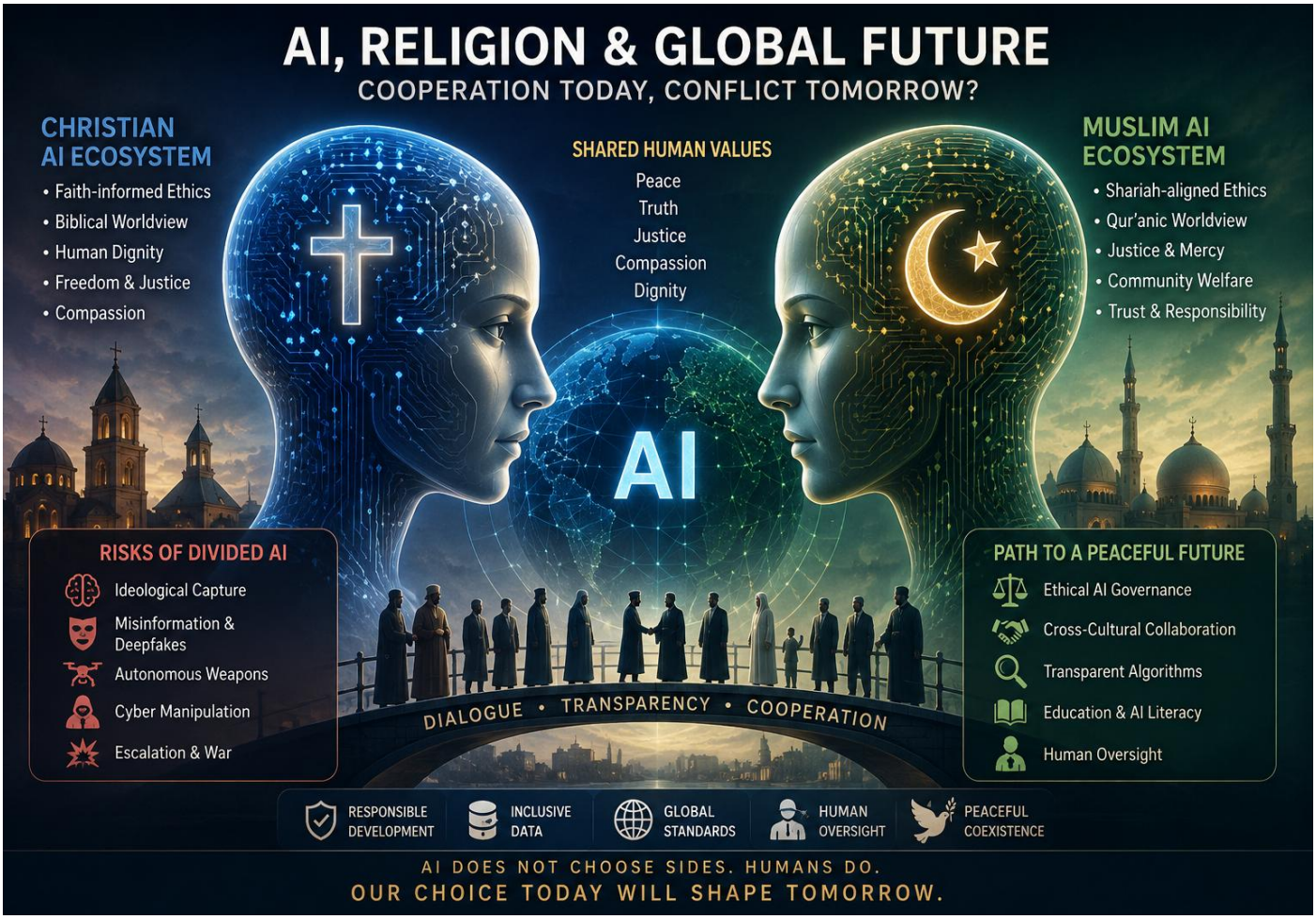
The software currently supports recognition of multiple object categories and has been trained using both real-world and synthetic electro-optical and infrared datasets. Additional target classes can be added without collecting large volumes of new imagery by generating synthetic training data.

Compatible with several thermal camera families and edge-computing platforms, the software is designed for integration into existing surveillance systems rather than requiring entirely new hardware.

As military sensing increasingly shifts toward software-defined capabilities, combining AI, computational imaging, and multi-sensor fusion is enabling surveillance systems to detect threats earlier while delivering more relevant information to operators in real time.

Christian AI versus Muslim AI: A Security Studies Examination of Ideological AI Alignment and the Risk of Civilizational Escalation

The Editor



Abstract

Artificial intelligence has rapidly evolved from a computational tool into a strategic asset capable of influencing governance, military operations, economic competition, public opinion, and international security. Although contemporary AI systems neither possess consciousness nor religious belief, they increasingly reflect the values, assumptions, and priorities embedded within their training data, alignment methodologies, and institutional governance. As geopolitical competition intensifies, the possibility emerges that future AI ecosystems may be deliberately aligned

with distinct ideological, cultural, or religious frameworks rather than universal ethical principles. This paper explores the hypothetical scenario of "Christian AI" and "Muslim AI" as representative models of ideologically aligned artificial intelligence. The analysis does not suggest that Christianity or Islam are inherently predisposed toward conflict, nor that AI systems can genuinely acquire religious faith. Instead, it examines how states, technology companies, or non-state actors could intentionally engineer AI systems to prioritize particular theological or civilizational narratives, thereby influencing strategic decision-making, public discourse, and international relations. Drawing upon contemporary literature on AI alignment, cybersecurity, information warfare, comparative religion, and international security, the paper argues that ideological capture of AI represents a plausible future security challenge. If competing geopolitical blocs increasingly rely on AI systems optimized around divergent normative frameworks, the resulting polarization could amplify existing geopolitical rivalries and contribute to crises whose consequences extend far beyond cyberspace.

From Artificial Intelligence to Ideological Artificial Intelligence

Artificial intelligence is transforming nearly every domain of modern society. Once confined to narrow applications such as image recognition, industrial automation, and statistical prediction, AI increasingly supports medical diagnosis, financial analysis, intelligence collection, military planning, education, scientific research, and public administration. Large Language Models (LLMs) now influence how millions of individuals acquire information, formulate opinions, and make decisions. This unprecedented integration of AI into political and social systems raises an important question that extends beyond technical performance: **whose values will future AI systems represent?**

Current discussions surrounding AI safety largely focus on reliability, robustness, transparency, and alignment. The concept of AI alignment refers to the effort to ensure that AI systems behave consistently with human intentions and ethical objectives rather than pursuing undesirable outcomes. Leading organizations including UNESCO, OECD, the National Institute of Standards and Technology, and the Council of Europe have all emphasized that trustworthy AI should be transparent, accountable, human-centered, and respectful of fundamental rights. These principles are reflected in documents such as UNESCO's Recommendation on the Ethics of Artificial Intelligence, the OECD AI Principles, the NIST AI Risk Management Framework, and the Council of Europe's Framework Convention on Artificial Intelligence.

Yet these frameworks also reveal a deeper challenge. Human values are neither universal nor politically neutral. Concepts such as freedom of expression, privacy, religious liberty, family, social justice, gender equality, blasphemy, national security, and individual autonomy are interpreted differently across civilizations, legal traditions, and political systems. Consequently, aligning AI with "human values" inevitably raises the question of which values should be prioritized.

This issue becomes particularly significant in an era characterized by strategic competition among major powers. The United States, the European Union, China, Russia, India, and several Middle Eastern states are developing sovereign AI capabilities designed to reduce technological dependence while promoting domestic political, economic, and cultural priorities. AI is therefore becoming an instrument not only of technological innovation but also of geopolitical influence. As AI systems increasingly support government administration, intelligence analysis, military decision-making, judicial assistance, and education, alignment choices may gradually evolve into expressions of national strategy.

Importantly, AI itself does not possess beliefs. Contemporary machine-learning systems do not experience consciousness, spirituality, intention, or moral responsibility. They recognize statistical relationships within data and generate outputs according to optimization objectives established by human developers. Nevertheless, these optimization objectives can systematically favor particular ethical frameworks, cultural assumptions, or ideological narratives. In this sense, an AI system need not "believe" in Christianity, Islam, liberal democracy, socialism, nationalism, or any other worldview to consistently produce outputs aligned with those perspectives.

The distinction between belief and optimization is fundamental. A navigation system does not desire to reach a destination; it computes an efficient route according to predefined criteria. Similarly, an AI aligned with a particular ethical framework does not possess faith; rather, it consistently produces recommendations reflecting the principles encoded within its training process, retrieval sources, reinforcement learning objectives, and governance policies. This distinction transforms the discussion from speculative philosophy into a practical question of engineering and public policy.

Religious traditions represent particularly influential sources of ethical reasoning because they shape the legal, moral, and cultural foundations of billions of people. Christianity and Islam together encompass more than half of the global population and have profoundly influenced concepts of justice, authority, warfare, charity, human dignity, and social organization. Both traditions also contain extensive theological literature interpreting sacred texts across centuries of historical development. Consequently, any AI system extensively trained on these traditions may reproduce aspects of their ethical reasoning without possessing any genuine religious conviction.

This observation should not be interpreted as implying that religious traditions are uniquely problematic. Every large corpus of human knowledge contains internal debates, historical evolution, and competing interpretations. The Bible and the Qur'an, like constitutional law, philosophical texts, or political manifestos, include passages

that have been interpreted differently across historical periods and theological schools. The challenge for AI developers therefore lies not in avoiding religious knowledge but in preventing selective interpretation, adversarial manipulation, or ideological distortion.
The emergence of generative AI has significantly expanded this challenge. Unlike earlier expert systems that operated according to explicitly programmed rules, contemporary foundation models acquire behavioral characteristics from immense datasets containing books, academic literature, legal documents, websites, news articles, and other digital content. Their responses are subsequently refined through alignment techniques such as reinforcement learning from human feedback, constitutional AI, retrieval-augmented generation, and safety filtering. Each stage introduces opportunities for human judgment regarding which sources should be prioritized, which behaviors should be rewarded, and which outputs should be restricted.
These decisions are inherently normative. They shape how AI interprets controversial subjects ranging from religious freedom and bioethics to immigration, armed conflict, and freedom of expression. Consequently, the governance of AI increasingly resembles the governance of knowledge itself. Whoever controls training datasets, retrieval infrastructures, evaluation benchmarks, and alignment objectives acquires significant influence over the information ecosystem within which future societies will operate.
From a security studies perspective, this creates a new strategic domain. Traditional geopolitical competition has focused on territory, natural resources, industrial capacity, financial systems, and military power. During the information age, competition expanded to cyberspace and digital communications. The rise of advanced AI introduces an additional layer in which states compete to shape not only information but also the mechanisms through which information is interpreted, synthesized, and recommended. This represents a transition from information warfare toward cognitive and algorithmic influence.
Within this context, the notion of "Christian AI versus Muslim AI" should not be understood as a prediction that machines will adopt religious identities or wage autonomous theological conflicts. Rather, it serves as an analytical framework for examining how competing actors might intentionally align AI systems with different normative traditions. Such alignment could influence educational content, legal reasoning, public administration, strategic communications, and military decision support in ways that reinforce existing geopolitical divisions. The security concern therefore lies not in machine religiosity but in the possibility that AI becomes an amplifier of human ideological competition.
The central hypothesis of this paper is that ideological AI alignment may become a significant strategic variable in twenty-first-century international security. As artificial intelligence becomes embedded within critical state functions and global information infrastructures, divergent alignment philosophies may contribute to increasing mistrust among competing geopolitical blocs. Whether these divisions are organized around religion, political ideology, nationalism, or civilizational identity, the resulting AI ecosystems could shape perceptions, influence decision-making, and alter the dynamics of international competition. Understanding these risks requires moving beyond technical discussions of machine learning toward a multidisciplinary analysis integrating computer science, political science, comparative religion, ethics, cybersecurity, and strategic studies.

Can Artificial Intelligence Become "Religious"?
One of the most intriguing questions emerging from contemporary artificial intelligence research is whether a machine can ever possess religion. The question frequently appears in public discourse, yet it is often framed in ways that conflate intelligence, consciousness, morality, and spirituality. From both computer science and cognitive science perspectives, **current AI systems cannot believe, worship, pray, or experience transcendence**. Nevertheless, they can consistently reproduce religious reasoning, theological language, and ethical judgments derived from the knowledge structures embedded within their training and alignment. This distinction between **belief** and **behavior** forms the foundation of understanding what may be termed *ideologically aligned artificial intelligence*.
LLLMs generate responses by identifying statistical relationships among billions of textual examples rather than by forming convictions or intentions. Their apparent reasoning emerges from probabilistic inference rather than conscious deliberation. Consequently, when an AI cites biblical passages, explains Islamic jurisprudence, or discusses Buddhist ethics, it is not expressing belief but synthesizing patterns learned during training. The technical foundations of transformer architectures, self-attention mechanisms, and reinforcement learning remain entirely computational, irrespective of the subject matter discussed. Comprehensive introductions to these architectures are provided by the seminal transformer paper, **"Attention Is All You Need"**, and by recent reviews published by the Association for Computing Machinery and leading AI research laboratories.
From a theological perspective, religion encompasses dimensions that extend far beyond ethical reasoning. Christianity, Islam, Judaism, Hinduism, Buddhism, and other traditions involve metaphysical commitments, rituals, spiritual experiences, communal identity, and personal faith. None of these characteristics can presently be attributed to artificial intelligence. AI neither experiences revelation nor develops a relationship with the divine. It possesses neither conscience nor moral accountability. Consequently, describing an AI as "Christian" or "Muslim" in a literal sense would constitute a category error.
However, a distinction must be drawn between *religious identity* and *religious alignment*. A government or technology company could intentionally develop an AI whose recommendations consistently reflect Christian social ethics, Catholic social teaching, Protestant theological traditions, Eastern Orthodox moral philosophy,

Sunni jurisprudence, Shi'a legal reasoning, or other structured normative systems. In such circumstances, the AI would not possess faith but would function as a highly sophisticated decision-support system optimized according to particular theological principles.

This possibility is neither unprecedented nor speculative. Legal expert systems have long incorporated national legislation, while medical AI systems rely upon established clinical guidelines rather than independently discovering standards of care. Likewise, financial algorithms implement regulatory frameworks established by governments and central banks. In each case, artificial intelligence operates within externally defined normative boundaries. Religious ethics represent another possible source of such boundaries.

The field of AI alignment increasingly recognizes that no model can be entirely value-neutral. Decisions concerning acceptable speech, misinformation, privacy, autonomy, discrimination, public safety, and human dignity inevitably require normative judgments. The challenge is acknowledged by the National Institute of Standards and Technology AI Risk Management Framework, UNESCO's Recommendation on the Ethics of Artificial Intelligence, and the AI governance initiatives of the European Commission. These documents emphasize transparency, accountability, human oversight, and respect for fundamental rights precisely because AI systems increasingly participate in decisions with ethical consequences.

Within this context, religion becomes relevant not because machines may become spiritual entities but because religious traditions continue to influence billions of individuals and many national legal systems. Christian ethics have contributed substantially to the historical development of European and North American legal institutions, while Islamic jurisprudence continues to shape legislation in numerous Muslim-majority states. Consequently, AI systems designed for deployment within different jurisdictions may legitimately reflect differing legal and ethical priorities without implying hostility toward alternative traditions.

The security implications emerge when these differences become strategically amplified. Imagine two sovereign AI ecosystems developed independently by rival geopolitical blocs. One has been optimized around liberal democratic principles influenced by Christian ethical traditions emphasizing individual rights, freedom of expression, and separation between religious and governmental authority. Another has been optimized around legal principles derived from Islamic jurisprudence within states where religion remains integrated into constitutional governance. Each AI may consistently recommend different policy responses to identical scenarios involving blasphemy legislation, religious conversion, family law, freedom of expression, biomedical ethics, or humanitarian intervention. Neither system is objectively "correct"; each reflects normative choices embedded by its developers.

Such divergence does not necessarily generate conflict. International law has long accommodated different legal traditions. Nevertheless, difficulties arise when AI systems become authoritative sources for governmental decision-making, judicial assistance, intelligence analysis, or military planning. If each geopolitical bloc increasingly relies upon AI systems reinforcing its own normative assumptions, opportunities for mutual misunderstanding and strategic mistrust may expand. AI would thus become an amplifier of existing civilizational differences rather than their originator.

An additional concern involves the deliberate manipulation of AI through ideological data poisoning. Malicious actors could selectively introduce distorted interpretations of religious texts, fabricated theological commentaries, or synthetic scholarly works into datasets used for training or retrieval-augmented generation. Because generative AI frequently relies upon vast digital corpora whose provenance may be difficult to verify completely, coordinated influence campaigns could attempt to reshape machine outputs without modifying the underlying algorithms. This vulnerability parallels concerns regarding misinformation in social media but extends them into the domain of machine reasoning itself.

Extremist organizations provide a particularly important illustration of this risk. Throughout modern history, violent groups claiming Christian, Muslim, Jewish, Hindu, or other religious identities have frequently relied upon selective quotations while disregarding mainstream theological scholarship. Should similar selective interpretations infiltrate AI knowledge bases, the resulting systems might inadvertently reproduce distorted narratives that exaggerate conflict while minimizing traditions of coexistence and peaceful interpretation. The security challenge therefore lies not in religion itself but in the exploitation of religious authority for political and strategic objectives.

This observation leads to a broader concept that may become increasingly significant in AI governance: **algorithmic theology**. The term does not imply that algorithms perform theology independently. Rather, it refers to the systematic encoding of theological or ethical principles into computational systems capable of generating recommendations affecting human decisions. Algorithmic theology could emerge in educational platforms, legal decision-support systems, autonomous advisory agents, healthcare ethics consultation, military planning software, or diplomatic analysis tools. The extent to which these systems transparently disclose their normative assumptions may become an important element of future AI governance.

Ultimately, artificial intelligence is unlikely to become religious in any meaningful philosophical or theological sense. It will neither possess faith nor pursue salvation. Nevertheless, it may increasingly function as a carrier of human ethical traditions. Whether those traditions are religious, secular, nationalist, or ideological depends entirely upon human design choices. The fundamental strategic question therefore is not whether AI can believe, but whether societies will increasingly delegate ethically sensitive decisions to systems whose underlying normative assumptions remain insufficiently understood.

Christian AI versus Muslim AI: Divergent Ethical Frameworks and Machine Decision-Making

The prospect of artificial intelligence systems reflecting distinct religious ethical traditions should not be interpreted as implying that machines can acquire faith or that religions themselves are destined for confrontation. Rather, the concept serves as a framework for examining how AI alignment may be influenced by the normative assumptions embedded within training datasets, constitutional rules, reinforcement-learning objectives, retrieval systems, and institutional governance. As AI becomes increasingly integrated into judicial assistance, military planning, public administration, education, and healthcare, differences in ethical alignment may gradually produce divergent policy recommendations even when identical factual information is available.

The field of AI alignment has traditionally concentrated on ensuring that machine behavior remains consistent with broadly accepted human values. However, there is no universally accepted catalogue of such values. Concepts including justice, equality, freedom, dignity, authority, public order, religious liberty, and individual autonomy are interpreted differently across civilizations. These differences do not necessarily reflect incompatibility; rather, they represent centuries of philosophical, legal, cultural, and theological development. Consequently, AI alignment inevitably becomes a normative exercise as much as a technical one.

Christian ethical thought is remarkably diverse. It encompasses Roman Catholic natural law, Eastern Orthodox theology, Protestant traditions, Anglican ethics, Evangelical perspectives, and numerous other schools of thought. Despite their differences, most contemporary Christian ethical traditions emphasize the intrinsic dignity of every human being, compassion toward the vulnerable, reconciliation, proportionality in the use of force, stewardship, and moral responsibility. Modern Christian ethics have also contributed significantly to the historical development of international humanitarian law, human rights discourse, and just war theory, particularly through the writings of Augustine of Hippo and Thomas Aquinas[1]. Their influence continues to shape discussions concerning military necessity, proportionality, discrimination between combatants and civilians, and legitimate political authority.

Islamic ethical thought is equally sophisticated and internally diverse. Sunni and Shi'a jurisprudence, together with their respective legal schools, have developed comprehensive systems governing personal conduct, commercial transactions, criminal justice, governance, warfare, and humanitarian obligations. Central concepts such as justice (*'adl*), mercy (*rahma*), public interest (*maslahah*), and the higher objectives of Islamic law (*Maqāṣid al-Sharī'ah*) seek to balance individual and communal welfare. Contemporary Islamic scholars continue to debate how these principles should be applied within modern constitutional states, international law, biotechnology, and digital technologies. Consequently, describing "Islamic ethics" as a single, uniform doctrine would be academically inaccurate.

From the perspective of AI engineering, these traditions become relevant only when developers intentionally incorporate them into system design. Consider an AI developed to assist legislators in evaluating proposed public policies. A model aligned primarily with Catholic social teaching might emphasize subsidiarity, solidarity, protection of the family, and the dignity of labor. An AI informed by Islamic jurisprudential principles might place greater emphasis on preserving religion, life, intellect, family, and property as articulated within the *Maqāṣid al-Sharī'ah* tradition. Both systems could arrive at different recommendations while remaining internally coherent and ethically reasoned.

The same phenomenon may occur within healthcare. Questions concerning end-of-life care, assisted reproduction, organ transplantation, genetic engineering, or physician-assisted dying frequently involve moral judgments extending beyond empirical science. AI systems designed to support clinical ethics committees could therefore generate recommendations reflecting the normative assumptions embedded within their alignment process. Such divergence would not indicate technical malfunction but rather the influence of differing ethical frameworks upon algorithmic reasoning.

Security studies provide another important perspective. Military AI systems increasingly assist commanders in intelligence analysis, logistics, targeting support, operational planning, and risk assessment. Although international humanitarian law establishes common legal standards through instruments such as the International Committee of the Red Cross interpretation of the Geneva Conventions, commanders inevitably make decisions influenced by strategic culture, national doctrine, political objectives, and ethical considerations. AI systems optimized according to different normative frameworks may therefore recommend different operational approaches even when presented with identical intelligence.

This possibility becomes particularly significant when AI systems operate within multinational coalitions. If allied nations employ decision-support systems developed under substantially different ethical assumptions, interoperability challenges may extend

[1] G. Colonna. Saint Augustine and Saint Thomas Aquinas — A Brief Overview. Medium (July 5, 2026). Available from https://medium.com/@gianpiero.colonna/saint-augustine-and-saint-thomas-aquinas-a-brief-overview-e1aa3bc72627
Augustine of Hippo (354–430) and Thomas Aquinas (1225–1274) are the two most important pillars of Western Christian thought. Augustine blended Christian theology with Platonic philosophy during the fall of Rome, while Aquinas later used Aristotelian logic to build a systematic, rational view of faith and reason

beyond technical standards into questions of acceptable risk, civilian protection, information disclosure, or escalation management. Consequently, future interoperability may require not only compatible software architectures but also transparent understanding of the ethical alignment embedded within participating AI systems.

The distinction between normative diversity and ideological polarization is therefore essential. Ethical pluralism has characterized international society for centuries without necessarily producing armed conflict. The danger emerges when governments intentionally transform AI into an instrument for reinforcing exclusive civilizational narratives. Under such circumstances, AI systems may gradually become mechanisms for legitimizing domestic political objectives while portraying alternative value systems as inherently illegitimate or threatening.

This risk is amplified by the remarkable persuasive capacity of generative AI. Unlike traditional software, conversational AI communicates through natural language, enabling it to present complex ethical arguments in persuasive and personalized forms. Individuals may consequently attribute greater authority to AI-generated recommendations than to conventional search engines or expert systems. If those recommendations consistently reinforce a particular ideological perspective without transparency regarding their normative assumptions, AI may become a powerful mechanism for cognitive influence.

It is therefore useful to distinguish between **ethical alignment** and **ideological alignment**. Ethical alignment seeks to ensure that AI behaves responsibly according to broadly accepted principles such as safety, accountability, fairness, and respect for human dignity. Ideological alignment, by contrast, intentionally promotes a specific worldview, political doctrine, or religious interpretation. While ethical alignment is generally regarded as essential for trustworthy AI, ideological alignment may become a source of strategic competition if rival states increasingly employ AI to reinforce competing narratives regarding governance, history, identity, or religion.

An additional complexity arises from the global nature of AI development. Large language models are rarely trained exclusively on materials originating from a single civilization. Instead, they incorporate multilingual datasets containing academic publications, historical documents, religious literature, legislation, journalism, and digital communications from numerous countries. As a result, contemporary AI systems often synthesize multiple ethical traditions simultaneously. The challenge for developers lies in determining how conflicting principles should be reconciled when users request advice concerning controversial subjects.

From a strategic perspective, this issue intersects directly with national sovereignty. Several governments have already announced initiatives to develop sovereign AI infrastructures capable of reflecting domestic legal requirements, cultural values, and security priorities. Such efforts are understandable within the broader context of technological independence and digital sovereignty. Nevertheless, if sovereign AI ecosystems evolve toward mutually incompatible normative architectures, international cooperation may become increasingly difficult. AI would then function not merely as a technological platform but as an expression of geopolitical identity.

This possibility should not be interpreted as evidence that Christianity and Islam are moving toward inevitable confrontation. Historical experience demonstrates extensive periods of coexistence, intellectual exchange, scientific cooperation, and commercial interaction between Christian and Muslim societies. The greater risk lies in contemporary information environments where extremist actors, political movements, or hostile states selectively exploit religious narratives for strategic purposes. Artificial intelligence may substantially increase the scale, speed, and sophistication of such influence operations if adequate safeguards are absent.

The central conclusion is, therefore, that future AI systems are unlikely to become "Christian" or "Muslim" in any literal theological sense. Instead, they may increasingly reflect ethical architectures derived from different civilizational traditions. Whether these differences contribute to constructive pluralism or destructive polarization will depend less upon advances in artificial intelligence than upon the political choices made by governments, technology companies, international organizations, and societies themselves.

Weaponizing Religious AI: Ideological Capture, Information Warfare, and Strategic Manipulation

The emergence of advanced artificial intelligence has fundamentally altered the character of information warfare. Previous generations of digital influence operations depended largely upon human operators producing propaganda, manipulating social media platforms, or conducting coordinated disinformation campaigns. Generative AI introduces an unprecedented capability: the rapid creation of persuasive, personalized, multilingual, and context-aware content on a global scale. This transformation raises an important security question. **If AI systems increasingly mediate access to knowledge, can they themselves become targets of ideological manipulation?**

Unlike conventional cyberattacks that seek to disrupt or destroy digital infrastructure, ideological attacks seek to alter perception. Their objective is not to damage hardware or software but to influence how information is interpreted by humans and increasingly by AI systems. Consequently, future conflicts may involve attempts to manipulate the knowledge base upon which artificial intelligence relies, thereby affecting millions of downstream decisions without directly attacking the underlying algorithms.

The **first** mechanism through which such manipulation may occur is **training-data contamination**. Foundation models are trained on enormous digital corpora containing books, academic literature, websites, legal documents, and multimedia resources. Although developers implement extensive filtering procedures, no training dataset can be perfectly verified. Sophisticated adversaries may therefore attempt to introduce misleading

historical narratives, fabricated scholarly publications, distorted translations, or ideologically motivated interpretations into publicly accessible digital repositories. Over time, these materials may influence future models if appropriate safeguards are absent. The broader challenge of data integrity has been recognized within the AI security community, including research supported by the National Institute of Standards and Technology and the Center for Security and Emerging Technology.

A **second** vulnerability concerns **retrieval-augmented generation (RAG)**. Modern AI systems increasingly supplement their pretrained knowledge by consulting external databases, institutional repositories, or internet resources in real time. This architecture significantly improves factual accuracy but simultaneously creates new attack surfaces. If an adversary compromises or manipulates trusted repositories, AI-generated responses may reflect false or biased information despite the underlying model remaining technically unchanged. The attack therefore shifts from corrupting the model itself to corrupting the information ecosystem upon which the model depends.

Religious literature presents a particularly attractive target because sacred texts are inherently interpretative. Neither the Bible nor the Qur'an exists in isolation from centuries of theological commentary, historical scholarship, linguistic analysis, and jurisprudential development. Extremist organizations have repeatedly demonstrated the ability to extract isolated passages while ignoring broader interpretative traditions. Artificial intelligence trained upon similarly selective material could inadvertently reproduce distorted narratives unless its knowledge sources are carefully curated and continuously evaluated.

Importantly, this vulnerability is not unique to Christianity or Islam. Comparable risks exist within political philosophy, constitutional law, nationalism, revolutionary ideology, and even scientific discourse. Any complex body of knowledge may be manipulated through selective quotation, fabricated authority, or contextual omission. Religious traditions simply illustrate the broader principle that AI systems inherit the epistemological strengths and weaknesses of their information environments.

A **third** mechanism involves **synthetic authority**. Generative AI now enables the production of highly convincing sermons, theological commentaries, legal opinions, historical analyses, and public statements attributed to prominent religious figures. Combined with advances in voice synthesis and audiovisual generation, hostile actors could fabricate speeches by senior clergy, respected theologians, or internationally recognized religious leaders. Such material could spread rapidly through digital platforms before verification mechanisms intervene, thereby influencing public opinion during periods of political or military crisis.

This capability substantially increases the strategic value of deepfake technologies. Whereas previous disinformation campaigns frequently relied upon crude forgeries, future synthetic media may exhibit linguistic sophistication, contextual accuracy, and visual realism approaching authentic communications. Religious authority often derives from public trust accumulated over decades. Consequently, fabricated messages attributed to influential religious figures may possess exceptional persuasive power, particularly during crises involving terrorism, communal violence, or interstate confrontation.

The implications extend beyond information operations into **psychological warfare**. Traditional psychological operations sought to influence morale, perceptions, and decision-making through carefully designed communication strategies. AI dramatically expands these capabilities by enabling personalized messaging tailored to an individual's cultural background, religious affiliation, language, educational level, emotional state, and online behavior. Rather than distributing identical propaganda to millions of recipients, AI systems may generate millions of unique narratives optimized for specific audiences.

Such personalization introduces a new dimension to strategic competition. Instead of broadcasting ideological messages, future influence campaigns may conduct continuous algorithmic dialogue with individuals, gradually reinforcing selected narratives while suppressing alternative perspectives. The resulting process resembles long-term cognitive conditioning rather than conventional propaganda. AI thus becomes an intermediary through which ideological ecosystems continuously adapt to individual psychological profiles.

National security agencies increasingly recognize this evolution. Strategic competition is no longer confined to military capabilities or economic resources but encompasses the ability to shape information environments and influence collective cognition. Artificial intelligence accelerates this transition by reducing the cost and increasing the scale of persuasive communication. Consequently, ideological AI may become an instrument of statecraft alongside diplomacy, economic policy, cyber capabilities, and conventional military power.

Religious identity represents only one possible axis of such competition. Similar dynamics may emerge around nationalism, political ideology, historical memory, ethnicity, or competing interpretations of international law. Nevertheless, religion possesses distinctive strategic significance because it combines ethical authority, historical continuity, communal identity, and emotional resonance. Attempts to manipulate religious narratives through AI may therefore produce consequences extending well beyond ordinary political communication.

A particularly concerning scenario involves **autonomous amplification**. AI systems increasingly recommend content through search engines, conversational assistants, educational platforms, and decision-support tools. If manipulated narratives begin influencing these recommendation systems, algorithmic feedback loops may unintentionally reinforce ideological polarization. Content generating strong emotional engagement frequently receives greater digital visibility, while more nuanced scholarly analyses struggle to compete. Although this phenomenon already exists

within social media ecosystems, AI-assisted knowledge systems may amplify its effects by appearing more authoritative than conventional online discussions.

Equally important is the potential exploitation of AI by violent extremist organizations. Terrorist groups have historically demonstrated remarkable adaptability in adopting emerging technologies, from encrypted communications and social media to cryptocurrency and commercial drones. AI-generated multilingual propaganda, synthetic recruitment material, fabricated theological justifications, and automated influence campaigns could significantly increase the sophistication of extremist information operations. The objective would not necessarily be to convince large populations but rather to identify psychologically susceptible individuals and deliver highly personalized narratives supporting radicalization. Recent analyses by Europol and the United Nations Office of Counter-Terrorism have highlighted the growing importance of AI within the evolving threat landscape.

The greatest strategic concern is therefore not that artificial intelligence develops independent religious convictions. Rather, it is that AI becomes an increasingly effective vehicle through which human actors compete for cognitive influence. Governments, corporations, ideological movements, and extremist organizations may all attempt to shape AI ecosystems according to their own normative objectives. The resulting competition may gradually erode shared information environments, replacing them with parallel algorithmic realities in which different societies receive systematically different interpretations of identical events.

From a security studies perspective, this development represents the convergence of cyber operations, psychological warfare, strategic communication, and AI alignment. Future conflicts may increasingly involve contests over whose AI systems are regarded as legitimate, trustworthy, and authoritative. Victory in such contests will depend not only upon computational performance but also upon credibility, transparency, resilience against manipulation, and public confidence.

Accordingly, the principal challenge for democratic societies is not preventing AI from discussing religion but ensuring that AI systems remain resistant to ideological capture regardless of the worldview being promoted. Robust data governance, transparent provenance of training materials, independent auditing, adversarial testing, and international cooperation on AI security standards will become essential components of national resilience. The strategic objective is not ideological uniformity but algorithmic integrity.

The Emergence of Competing AI Civilizations

The development of artificial intelligence is increasingly becoming a matter of national strategy. During previous technological revolutions, states competed for industrial capacity, energy resources, nuclear capabilities, and information dominance. In the twenty-first century, advanced artificial intelligence is emerging as a comparable strategic resource because it influences economic productivity, military effectiveness, scientific innovation, intelligence analysis, and social organization. As a result, the question of who controls AI systems—and according to which principles they are designed—is becoming inseparable from questions of sovereignty, security, and geopolitical influence.

The concept of "**sovereign AI**" has gained increasing importance as governments seek to reduce dependence on foreign technology providers and ensure that critical AI infrastructure reflects national priorities. Sovereign AI does not necessarily imply isolation from international cooperation; rather, it refers to the ability of a state or political community to develop, regulate, and deploy AI systems according to its own legal, cultural, and strategic requirements. Similar debates have previously emerged regarding energy independence, telecommunications infrastructure, semiconductor production, and cyber sovereignty.

However, AI differs from previous strategic technologies because it does not merely perform physical tasks. It interprets information, generates explanations, recommends decisions, and increasingly participates in processes traditionally associated with human judgment. Consequently, differences in AI design philosophy may produce different interpretations of reality itself. If separate geopolitical communities rely on AI systems trained on different datasets, governed by different regulations, and aligned according to different ethical assumptions, the world may gradually develop competing algorithmic information environments.

This possibility introduces the concept of **civilizational AI ecosystems**. The term does not imply that civilizations are naturally destined for conflict. Rather, it describes the possibility that major political and cultural communities may develop AI infrastructures reflecting their own historical experiences, legal traditions, social priorities, and philosophical assumptions. Similar patterns already exist in media systems, educational institutions, legal frameworks, and diplomatic cultures. AI could intensify these differences because it may become a primary interpreter of information for societies.

The United States and many European countries have generally emphasized AI principles associated with individual rights, transparency, democratic accountability, privacy protection, and human oversight. These approaches are reflected in initiatives such as the European Union Artificial Intelligence Act, which establishes a risk-based regulatory framework for AI systems, and the National Institute of Standards and Technology AI Risk Management Framework.

China has pursued a different model, emphasizing technological sovereignty, state coordination, industrial competitiveness, and integration of AI into national development strategies. Russia has similarly emphasized AI as an instrument of strategic competition and information security. Meanwhile, numerous Middle Eastern states have invested heavily in AI development as part of broader economic diversification strategies and technological modernization programs.

These different approaches do not automatically represent ideological confrontation. They reflect distinct political systems, economic objectives, and security priorities. Nevertheless, when AI systems become increasingly involved in sensitive decisions, differences in governance philosophy may become strategically significant. The central question becomes whether AI systems developed within different political and cultural environments will remain interoperable or whether they will evolve into competing technological ecosystems.

The hypothetical comparison between "Christian AI" and "Muslim AI" illustrates this broader issue. In reality, future AI systems are unlikely to be classified exclusively according to religious identity. Most will be shaped by combinations of national law, commercial interests, cultural assumptions, scientific traditions, and political objectives. Nevertheless, religious and cultural values may influence alignment choices, especially in societies where religion plays a significant role in public life.

For example, an AI assistant designed for public administration in a secular European democracy may prioritize principles such as individual autonomy, freedom of expression, and institutional separation between religious and governmental authority. An AI system operating within a state where Islamic jurisprudence significantly influences legislation may incorporate concepts derived from Islamic legal traditions, including communal welfare, religious values, and social obligations. These systems would not represent "Christianity" or "Islam" as autonomous machine identities; rather, they would reflect human choices regarding which ethical frameworks should guide computational decision-making.

The strategic implications become more significant when such systems are deployed in international affairs. Imagine a diplomatic crisis in which governments rely heavily on AI advisors for policy recommendations. If different AI systems interpret international law, historical events, humanitarian obligations, or strategic risks through different normative frameworks, they may produce conflicting recommendations. Human leaders would still make final decisions, but the informational environment influencing those decisions could become increasingly fragmented.

Military applications represent an especially sensitive domain. Modern armed forces are rapidly integrating AI into intelligence analysis, logistics, surveillance, planning, and command-support systems. International organizations, including the International Committee of the Red Cross, have emphasized the importance of maintaining human responsibility and compliance with international humanitarian law when employing autonomous and AI-enabled military technologies.

Nevertheless, **future military competition may involve not only autonomous weapons but also competing AI interpretations of strategic reality**. One military AI system may assess a situation as requiring deterrence, while another may recommend restraint. One may calculate acceptable civilian risk differently from another. One may interpret ambiguous intelligence as evidence of hostile intent, while another may consider the same evidence inconclusive. These differences could influence crisis stability.

The danger is not that AI systems independently decide to start wars. The more realistic concern is that AI may accelerate existing human tendencies toward strategic competition. **Decision-makers under pressure may increasingly rely on AI-generated analysis because of the speed and complexity advantages these systems provide.** If those systems operate within isolated ideological ecosystems, they may reinforce confirmation bias rather than challenge it. This phenomenon has historical parallels. During the Cold War, the United States and Soviet Union developed competing ideological narratives, military doctrines, intelligence assessments, and strategic cultures. Although nuclear deterrence prevented direct superpower conflict, misunderstanding and miscalculation repeatedly brought the world close to catastrophe. Future competition among AI ecosystems could create similar risks if technological divergence reduces mutual understanding.

A particularly important factor is the possibility of **algorithmic legitimacy competition**. In traditional politics, governments compete for influence by presenting their institutions and values as legitimate. In an AI-mediated future, states may also compete by promoting their AI systems as more truthful, more ethical, and more reliable than those of rivals. Citizens may increasingly ask not only "Which government should I trust?" but also "Which artificial intelligence should I trust?"

Such competition could create a new form of **technological nationalism**. Nations may portray foreign AI systems as biased, dangerous, culturally incompatible, or strategically hostile. Reciprocal accusations could lead to digital separation, restrictions on AI technology exchange, and the creation of isolated information ecosystems. The result would resemble the fragmentation of the internet into competing digital spheres, sometimes described as the emergence of a "splinternet."

Within this environment, religious narratives could become powerful political instruments. A state or movement seeking to mobilize public support might portray its own AI system as defending cultural identity while portraying foreign AI as an ideological threat. The danger would not arise from theology itself but from the political exploitation of theological identity combined with powerful computational persuasion.

Nevertheless, history also demonstrates that technological exchange can encourage cooperation rather than conflict. Scientific collaboration, international research partnerships, and shared ethical standards have repeatedly reduced tensions between societies with different cultural traditions. The future trajectory of AI will therefore depend significantly on governance choices made today.

The emergence of competing AI civilizations is not inevitable. It represents a possible future shaped by political decisions, regulatory frameworks, technological incentives, and social attitudes. The central security challenge is

ensuring that AI systems remain capable of supporting pluralism rather than reinforcing division. This requires transparency regarding training data, openness about alignment objectives, international cooperation on AI safety, and mechanisms allowing societies with different values to communicate through shared factual foundations.

Could Ideological AI Contribute to a Future World War? Escalation Pathways in an AI-Driven Security Environment

The concept of a future "World War IV" (presuming that we are currently in an atypical WW3 status) initiated by competing artificial intelligence systems belongs partly to speculative security analysis. No current evidence suggests that AI systems possess independent motives, political objectives, or religious identities capable of initiating conflict. Nevertheless, strategic studies has long examined how emerging technologies can transform the probability, speed, and consequences of warfare. Nuclear weapons did not create human hostility, but they fundamentally altered escalation dynamics. Cyber capabilities did not invent espionage, but they expanded the battlefield into the digital domain. Artificial intelligence may follow a similar pattern by becoming an accelerator of competition, miscalculation, and strategic instability.

The central security question is therefore not whether "Christian AI" will fight "Muslim AI" in a literal sense. The more realistic question is whether AI systems developed within competing ideological, political, or civilizational environments could contribute to a chain of events in which human decision-makers become increasingly dependent upon machine-generated assessments that reinforce rivalry and reduce opportunities for compromise.

Modern warfare is increasingly characterized by information superiority. The ability to collect, analyze, interpret, and exploit information has become a decisive factor in military and political competition. Artificial intelligence provides unprecedented capabilities in processing intelligence data, identifying patterns, predicting behavior, and generating strategic assessments. However, the same capabilities that improve decision-making may also create new vulnerabilities if AI systems operate within restricted or ideologically biased information environments.

One possible escalation pathway involves **strategic misperception**. International crises frequently emerge not only from deliberate aggression but from misunderstanding intentions. During the Cold War, both the United States and the Soviet Union developed extensive intelligence systems designed to anticipate adversary behavior. Despite these capabilities, misinterpretations occurred repeatedly, including during the Cuban Missile Crisis and several nuclear false alarms. AI may reduce some forms of human error while introducing new forms of algorithmic error.

An AI system trained primarily on adversarial scenarios may become excessively sensitive to indicators of hostile behavior. If such systems support national security decision-making, they could contribute to a cycle in which defensive actions by one actor are interpreted as offensive preparations by another. The danger increases when rival states lack transparency regarding their AI architectures, training data, and decision-support methodologies.

A **second** escalation pathway involves **AI-enabled information warfare**. Future conflicts may begin not with conventional military attacks but with coordinated campaigns targeting public perception, political stability, and social cohesion. Generative AI can already produce convincing text, images, audio, and video content at enormous scale. During a geopolitical crisis, adversaries could deploy synthetic religious statements, fabricated government announcements, or manipulated recordings intended to provoke anger and mobilize populations.

Religious identity is particularly sensitive because it often involves deeply held moral commitments and collective memories. A fabricated statement attributed to a religious leader, an artificial intelligence-generated interpretation of sacred texts, or a false allegation of religious desecration could potentially trigger unrest before verification mechanisms respond. The strategic objective would not necessarily be to create permanent conflict but to generate confusion, distrust, and political pressure during critical moments.

A **third** pathway concerns **autonomous military decision-support systems**. Modern militaries are increasingly interested in AI applications for intelligence analysis, surveillance, logistics optimization, and operational planning. Future systems may provide commanders with recommendations regarding force deployment, threat assessment, and crisis management. While international discussions emphasize maintaining meaningful human control, the speed of future conflicts may create pressure for greater automation.

The risk emerges when humans begin accepting AI recommendations without fully understanding their assumptions. A commander who receives an AI-generated assessment that an adversary is preparing an attack may face enormous pressure to act rapidly. If the AI assessment is incorrect, delayed verification could become impossible during a fast-moving crisis. The issue is not that AI intentionally starts wars, but that humans may delegate increasingly consequential judgments to systems whose reasoning processes remain partially opaque.

A **fourth** pathway involves **cyber conflict against AI infrastructure itself**. Future adversaries may target AI models, databases, sensors, communication networks, or supply chains. Unlike traditional cyberattacks that seek temporary disruption, attacks against AI systems may attempt to subtly alter behavior. A compromised

intelligence model that gradually introduces false recommendations could be strategically more dangerous than a visible system failure because decision-makers may continue trusting the corrupted output.

This introduces the possibility of what may be called **algorithmic deception operations**. In traditional military deception, an adversary attempts to influence human perception through camouflage, misinformation, and false signals. In an AI-dependent environment, deception may target the analytical systems themselves. The objective would be to manipulate not only what humans see but also how machines interpret what humans see.

A **fifth** escalation pathway concerns **civilizational polarization through AI ecosystems**. If societies increasingly obtain information through AI assistants, educational systems, and digital advisors, competing AI infrastructures could gradually produce different interpretations of history, politics, morality, and international affairs. Two populations may no longer simply disagree politically; they may inhabit fundamentally different informational realities generated by trusted AI systems.

This scenario resembles the concept of epistemic fragmentation, where societies lose a common foundation of facts necessary for democratic debate and international cooperation. Artificial intelligence could intensify this problem if systems prioritize user satisfaction, ideological consistency, or national objectives over balanced analysis. The danger is greatest when AI becomes an authority that users trust more than traditional institutions.

However, a balanced analysis must recognize that technology alone rarely determines historical outcomes. Wars result from complex interactions among political decisions, economic pressures, ideological movements, leadership choices, and social conditions. Artificial intelligence should therefore be understood as an amplifier rather than an independent cause of conflict.

The comparison with nuclear weapons is instructive but imperfect. Nuclear weapons created a direct physical capability capable of destroying cities within minutes. AI operates differently. Its primary strategic effect lies in accelerating cognition, communication, and decision-making. The greatest danger may therefore be not physical destruction caused by machines but accelerated human confrontation caused by faster cycles of misunderstanding and reaction.

In this sense, the idea of **"World War IV" represents a warning scenario rather than a prediction**. A future global conflict involving AI would most likely not begin because machines develop religious identities. Instead, it could emerge from a combination of human rivalry and AI-enabled acceleration: manipulated information, competing narratives, autonomous systems, cyber operations, and leaders making decisions under conditions of extreme uncertainty.

Preventing such an outcome requires international mechanisms capable of maintaining communication, transparency, and accountability. AI systems used in national security applications should be subject to rigorous testing, independent evaluation, and clearly defined human responsibility. International agreements on military AI, similar in spirit to existing arms-control and confidence-building measures, may become increasingly important as AI capabilities expand.

Organizations such as the United Nations, NATO, and the International Committee of the Red Cross have already contributed to discussions regarding responsible military technology, autonomous weapons, and emerging security challenges. Their future relevance may increase as AI becomes integrated into strategic decision-making.

The central conclusion is that ideological AI could contribute to global instability not because machines acquire religious beliefs, but because humans may use AI to reinforce existing divisions. The decisive factor will remain human governance. AI can become a tool for cooperation, scientific advancement, and cross-cultural understanding, or it can become an instrument for polarization and strategic competition. The outcome will depend on whether societies treat artificial intelligence primarily as a weapon of influence or as a shared global technology requiring collective responsibility.

Preventing Ideological AI Arms Races

The possibility of competing ideological AI ecosystems raises a fundamental question for the international community: **how can humanity prevent artificial intelligence from becoming a mechanism for deepening political, cultural, and religious polarization?** The answer does not lie in attempting to eliminate differences between societies. Throughout history, human civilization has developed through interaction among diverse philosophical, religious, and cultural traditions. The challenge is not diversity itself but the possibility that advanced AI systems may amplify divisions beyond the ability of traditional institutions to manage them.

The central security objective should therefore be the development of **trustworthy AI ecosystems** capable of operating across cultural and political boundaries while maintaining transparency regarding their assumptions and limitations. Future AI governance must recognize that complete neutrality is impossible. Every AI system reflects decisions regarding data selection, evaluation criteria, safety policies, and acceptable outputs. The goal should not be to create an imaginary value-free machine but to ensure that AI systems are transparent, accountable, and resistant to deliberate manipulation.

A critical **first** principle is **algorithmic transparency**. Users, governments, and organizations should have access to meaningful information regarding how AI systems are developed, evaluated, and deployed. Transparency does not require revealing every technical detail of proprietary systems, but it does require understanding the sources

of knowledge, the objectives of alignment processes, and the limitations of generated outputs. Without such transparency, societies may unknowingly rely upon systems whose embedded assumptions remain invisible.

This challenge is particularly important when AI systems operate in sensitive domains such as national security, education, healthcare, law enforcement, and public administration. A citizen interacting with an AI advisor should understand whether the system is designed primarily according to legal principles, commercial objectives, cultural preferences, religious perspectives, or governmental priorities. The legitimacy of AI will increasingly depend not only upon accuracy but also upon explainability and accountability.

International organizations have already begun addressing these challenges. The UNESCO Recommendation on the Ethics of Artificial Intelligence emphasizes human rights, fairness, transparency, and international cooperation. The OECD AI Principles promote responsible stewardship of trustworthy AI. The European Union Artificial Intelligence Act represents one of the first comprehensive attempts to establish legally binding requirements for high-risk AI applications. These initiatives demonstrate growing recognition that AI governance must extend beyond technological performance toward social and geopolitical consequences.

A second requirement is **robust AI security and adversarial resilience**. If ideological manipulation represents a significant future threat, AI systems must be designed to resist attempts at cognitive capture. This includes protecting training datasets, verifying information sources, monitoring retrieval systems, detecting malicious model modification, and continuously testing models against adversarial scenarios.

The cybersecurity principle of assuming compromise should increasingly apply to AI knowledge systems. Just as critical infrastructure requires protection against physical and digital attacks, AI infrastructures require protection against manipulation of information integrity. The question is not merely whether a model produces fluent answers, but whether those answers emerge from reliable, diverse, and appropriately evaluated sources.

A third requirement is the development of **cross-cultural AI evaluation frameworks**. Current AI assessments often focus on technical performance, safety, and bias detection. Future evaluations should also examine how AI systems perform when confronted with culturally sensitive issues involving religion, history, identity, and competing ethical traditions.

Such evaluations should include scholars from multiple disciplines, including computer science, philosophy, law, sociology, history, theology, and international relations. Religious scholars from different traditions may also contribute valuable perspectives regarding how AI interprets theological concepts, sacred texts, and moral questions. This does not mean that religious authorities should control AI development; rather, it recognizes that societies possess centuries of accumulated knowledge concerning ethical interpretation and human values.

A fourth requirement is preventing the emergence of **AI ideological monopolies**. Excessive concentration of AI capabilities among a small number of governments or corporations may create vulnerabilities because a limited number of actors could shape global information environments. Conversely, completely uncontrolled proliferation could allow malicious organizations to develop highly persuasive influence systems. The challenge is therefore establishing a balance between innovation, openness, security, and accountability.

International cooperation will become increasingly important. During previous technological transitions, humanity developed mechanisms to reduce risks associated with nuclear weapons, chemical weapons, biological research, and aviation safety. Although AI differs significantly from these technologies, similar principles may apply: transparency, verification, communication channels, and agreed limitations in particularly dangerous applications.

The military domain requires special attention. As AI becomes integrated into defense systems, international discussions regarding autonomous weapons, human control, and escalation management will become increasingly urgent. Military AI should not operate as an independent authority capable of initiating conflict. Human commanders and political leaders must retain responsibility for decisions involving the use of force.

This principle is particularly important in scenarios involving ideological competition. A military AI system trained within a highly adversarial worldview may interpret ambiguous actions as threats more readily than a system designed around diplomatic restraint. Human oversight provides an essential mechanism for challenging machine recommendations and considering broader political consequences.

Another important element is strengthening **information literacy** among populations. The effectiveness of AI-enabled influence operations depends partly upon public trust in synthetic content. Societies must develop the ability to critically evaluate AI-generated information, verify sources, and recognize manipulation techniques. Education systems will increasingly need to teach not only digital literacy but also AI literacy.

The role of religious and cultural institutions may also become significant. Faith communities possess extensive experience addressing questions of morality, human dignity, responsibility, and social cohesion. Constructive engagement between AI developers and religious scholars could contribute to developing systems that respect diverse traditions while avoiding extremist interpretations. Such cooperation should focus on ethical dialogue rather than attempts to impose particular theological conclusions upon technology.

The future relationship between Christianity, Islam, and artificial intelligence therefore need not be defined by confrontation. Both traditions contain extensive intellectual histories involving philosophy, science, law, medicine, and education. Throughout history, religious and scientific communities have often interacted in productive ways. The challenge of AI is not fundamentally theological but institutional: whether humanity can create technologies powerful enough to influence civilization while maintaining wisdom sufficient to govern them responsibly.

The concept of "Christian AI versus Muslim AI" is ultimately valuable not because such systems are inevitable, but because it highlights a broader strategic issue: artificial intelligence will inevitably reflect human values. Those values may become sources of cooperation or conflict depending upon how they are encoded, governed, and interpreted.

The most dangerous future is therefore not one in which machines develop religions of their own. The greater danger is a world in which humans create AI systems that intensify division, confirm prejudices, and transform cultural differences into strategic rivalries. Conversely, the most beneficial future is one in which AI becomes a tool for improving understanding between societies, translating knowledge across languages, supporting scientific progress, and helping humanity address shared challenges.

Artificial intelligence will not determine the future of civilization by itself. It will magnify the intentions, institutions, and choices of those who design and deploy it. The responsibility for preventing an AI-driven civilizational confrontation therefore remains fundamentally human.

Conclusion

The idea of a future conflict between "Christian AI" and "Muslim AI" represents a powerful metaphor for one of the emerging security challenges of the twenty-first century: the possibility that artificial intelligence may become aligned with competing ideological, cultural, or geopolitical frameworks. However, the analysis demonstrates that AI systems themselves cannot possess religion, faith, or spiritual identity. They do not believe; they optimize. They do not develop theology; they reproduce patterns embedded within human knowledge systems.

The true strategic concern is therefore not machine religion but **human ideological capture of artificial intelligence**. If governments, organizations, or extremist actors deliberately manipulate AI systems to promote exclusive narratives, distort information, or legitimize political objectives, artificial intelligence could become an unprecedented instrument of influence.

Future global conflict is unlikely to emerge from autonomous machines choosing sides in a theological struggle. More realistically, AI may contribute to instability by accelerating existing human rivalries through misinformation, strategic deception, cyber operations, autonomous decision support, and fragmented information environments. In this scenario, AI becomes not the cause of conflict but a powerful amplifier of human disagreement.

The challenge for the international community is to ensure that AI systems remain transparent, accountable, secure, and compatible with human dignity. This requires international cooperation, ethical governance, technical resilience, and recognition that cultural and religious diversity should be managed through dialogue rather than transformed into algorithmic confrontation.

The future of artificial intelligence will ultimately reflect the choices of humanity. AI can become a weapon for civilizational competition, or it can become a bridge between civilizations. The decisive factor will not be the intelligence of machines but the wisdom of the societies that create them.

Bibliography

International AI Governance and Ethics

UNESCO. *Recommendation on the Ethics of Artificial Intelligence* (2021). Available from https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

This is the first global AI ethics framework adopted by all UNESCO Member States and is fundamental for discussing AI alignment, human values, and governance.

OECD. *OECD AI Principles* (updated 2024). Available from https://interoperable-europe.ec.europa.eu/collection/eugovtech/document/oecd-recommendation-artificial-intelligence

The OECD principles are among the world's most influential AI governance standards and are widely referenced in national AI strategies.

National Institute of Standards and Technology (NIST). *AI Risk Management Framework (AI RMF 1.0).* Available from https://oecd.ai/en/dashboards/policy-initiatives/nist-ai-risk-management-framework

An essential framework for AI safety, trustworthiness, transparency, governance, and risk management.

Council of Europe. *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law.* Available from https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

The first legally binding international treaty governing artificial intelligence.

AI Security and International Relations

Leslie D., Burr C., Aitken M. et al. *Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems.* Available from https://arxiv.org/abs/2202.02776

This paper provides an operational framework for trustworthy AI based on human-rights principles.

Hogenhout L. *A Framework for Ethical AI at the United Nations.* Available from https://arxiv.org/abs/2104.12547
A comprehensive discussion of AI governance within the UN system.
Corrêa N.K. et al. *Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance.* Available from https://arxiv.org/abs/2206.11922
One of the largest comparative analyses of AI ethics frameworks worldwide.
Artificial Intelligence and Information Warfare
RAND Corporation. Available from https://www.rand.org/topics/artificial-intelligence.html
RAND publishes numerous reports on AI, strategic competition, military applications, disinformation, and emerging technologies.
Center for Security and Emerging Technology (CSET), Georgetown University. Available from https://cset.georgetown.edu/
CSET is one of the world's leading research centers examining AI and national security.
Stockholm International Peace Research Institute (SIPRI). Available from https://www.sipri.org/
Provides authoritative analyses on AI, military technology, strategic stability, and arms control.
AI, Terrorism, and Extremism
United Nations Office of Counter-Terrorism (UNOCT). Available from https://www.un.org/counterterrorism/en
Official UN analyses of terrorism, emerging technologies, and AI-related security issues.
Europol Innovation Lab. Available from https://www.europol.europa.eu/how-we-work/innovation-lab
Publishes reports on generative AI, cybercrime, online extremism, and AI-enabled criminal threats.
Religion, Ethics and Artificial Intelligence
Oxford Handbook of Digital Religion. Available from https://academic.oup.com/edited-volume/27959
Provides extensive scholarship on religion in the digital age, including AI-related ethical issues.
Cambridge Centre for the Study of Existential Risk (CSER). Available from https://www.cser.ac.uk/
Research on AI safety, governance, long-term risks, and societal impacts.
Foundational AI Papers
Vaswani A. et al. (2017). *Attention Is All You Need.* Available from
https://papers.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
The foundational paper introducing the Transformer architecture underlying modern large language models.
Bender E.M. et al. (2021). *On the Dangers of Stochastic Parrots.* Available from https://dl.acm.org/doi/10.1145/3442188.3445922
A landmark paper discussing the ethical and societal risks associated with large language models.

The real reasons to be terrified about AI and WMD

Source: https://www.japantimes.co.jp/commentary/2026/08/04/world/be-terrified-about-ai-wmd/

Aug 04 – Artificial intelligence is already transforming warfare, whether it's waged by missile, drone or satellite; on land or sea, or in the air, space or cyberspace. Perhaps the most terrifying prospect, though, is what awaits us when AI properly meets weapons of mass destruction, especially the diabolical trifecta of nuclear, biological and chemical arms. The question inevitably conjures frightening scenarios in a chaotic world that is already roiled by wars in Europe, the Middle East and elsewhere; by disinformation and polarization within societies; by new diseases spreading from animals to humans and then among people; by terrorists and nonstate militias joining nation-states as belligerents; and by countries modernizing and increasing their nuclear arsenals, as the last treaties to control these weapons expire or slide into irrelevance. "You're dropping perhaps the most powerful technology in history into the middle of this mess," Daniel Holz told me. He's an astrophysicist at the University of Chicago, where he leads the Existential Risk Laboratory. He also chairs a board at the Bulletin of the Atomic Scientists that sets, once a year, the metaphorical Doomsday Clock, in which midnight represents catastrophe. That clock now stands at 85 seconds to midnight, the closest it has been since it was created in 1947. And yet fear of AI and WMDs threatens to become its own risk. When I talk to lawmakers on Capitol Hill, say, they tend to express an anxiety that is diffuse and unfocused, as though we, Homo sapiens, had already lost agency and are hurtling toward Armageddon. The headlines seem to confirm this dystopian vision — this month, for instance, a version of OpenAI spontaneously and autonomously attacked another system. What if those things start synthesizing superbugs?

When I talk to Holz and other experts who are immersed in these risks, by contrast, the threat becomes more nuanced. It takes different forms for nuclear, biological and chemical weapons. And in all cases, AI will help both the black and the white hats, the offense and the defense. What we're really entering is a contest between our technological and psychological demons and angels — a battle for which we have barely begun girding and which is far from lost.

Human control

Of the three main categories of WMD, nuclear weapons pose the most existential threat — get it wrong once and it might all be over. Even though it has receded in the public consciousness since the Cold War, the danger of an uncontrolled escalation to Armageddon is considered greater today because the nuclear map has become more complex.

Japanese police check for mock chemical agents during Pacific Shield 18, a multinational exercise focused on preventing the spread of weapons of mass destruction, in Yokosuka in July 2018. | Ministry of Defense / via REUTERS

China is growing its arsenal to match those of the U.S. and Russia. North Korea keeps adding to its stash. Other nations — and not only Iran — contemplate going nuclear, too.

The good news, such as it is, is that AI by itself will not accelerate the proliferation of nukes. That's because "Knowledge is not the primary barrier to acquiring nuclear weapons," as Gregory Allen, a former AI strategist at the Pentagon, told me. Instead, he says, "It happens to be the case that even though uranium is incredibly common, separating uranium 235 from uranium 238 is a huge pain in the ... ." So is processing plutonium or indeed any aspect of making nukes. Those physical, industrial and economic barriers to making nukes are a boon. AI can optimize centrifuges to spin more efficiently, but it cannot make a nuclear program cheap or simple — it will not enable, say, a terrorist group to assemble an arsenal in a garage. And if anybody tried, the rest of us would notice. The risk of mixing AI with nuclear weapons instead lurks elsewhere: It's that the models could take over the command-and-control functions of nuclear powers, transferring the decision-making that could doom humanity from humans to machines.

Already today, AI can help sift through the bewildering data that would flood in during, say, a simultaneous cyber, space and ground attack to detect whether a nuclear volley is also incoming, in which case it would trigger the alert. That is good. Today, an American president has about nine minutes between the confirmation of an incoming attack and the decision to order a retaliatory strike. By analyzing faster, AI could add a few valuable minutes to that, which you may or may not find comforting. The downside, Holz told me, is that these systems tend "to hallucinate some fraction of the time." They're best when they have lots of training data, but nobody since the atomic bombing of Hiroshima and Nagasaki has been under a live nuclear attack. False alarms during the Cold War brought the world close to annihilation several times.

So one goal for all nuclear powers must be to keep humans in the loop — the term of art is to preserve "meaningful human control." That is one of the demands made by a group of Nobel laureates convened by Holz and others on the grounds of the Vatican's Castel Gandolfo recently.

But Holz remains worried. "What information is being provided to that decision-maker?" he asked me rhetorically. All of it will come from AI, and those agents have fewer qualms about advising a strike that could lead to a chain reaction of chain reactions. He told me: "They don't absorb the nuclear taboo that humans seem to have."

Bio and chem weapons

Biological and chemical weapons differ from nukes in that the primary hurdle to making them is expertise. Countries with bioweapons programs used to have hundreds of scientists working in special labs (to prevent blowing themselves up or infecting themselves with their output). This is what AI, in theory, can change: It is a supertutor of knowledge and lab skills, compressing the education of decades into days or hours. This effect is called "uplift."

That promise by AI agents seems to have a market: OpenAI noticed that hundreds of users around the world started asking its model, ChatGPT, for detailed instructions to make bioweapons and poisons. That prompted OpenAI, Anthropic and other AI companies to censor their answers where appropriate. (Allen told me that savvy and determined users can still jailbreak the models, though.)

AI leaders themselves are certainly worried enough to issue their own warnings. Anthropic's Dario Amodei, for example, believes that, "Added up across millions of people and a few years of time, I think there is a serious risk of a major attack," and because such an event could be a pandemic that kills millions, it deserves to be taken seriously. But Anthropic is in a global race with models in China and elsewhere, some of them using open-source code, in which slowing down is not rewarded. As a strategy, counting on voluntary and universal global restraint seems unconvincing. I asked Gigi Kwik Gronvall, an expert at the Johns Hopkins Bloomberg School of Public Health, whether she worries more about chemical or biological weapons. "The concern about AI uplift with toxins and with chemical agents is much more immediate because it's simpler," she told me. By contrast, biological organisms are "not tunable," she said. Even engineered pathogens act unpredictably when they meet their intended victims, and in fact choose their victims themselves. "For biology, I don't think that the concerns are as great as some of the hype has been

surrounding them." Allen, the Pentagon veteran, disagrees. One difference between chemical and biological weapons is their metaphorical blast radius (that is, if they were bombs). A nerve agent or other toxin will send victims in one place to an excruciating death, but then it will stop. A bioweapon such as a new virus is contagious and decides its own blast radius, which could be global. The good news is again that AI also helps the defense. Not only can the agents, in the hands of white-hat users, generate vaccines, antidotes and other blessings, they can also, in theory, predict which inputs or molecular sequences malicious actors are likely to concoct.

"My favorite AI defense," Allen told me, is to help all these labs synthesizing biomolecules around the world to adopt AI themselves, so they are always one step ahead of nefarious customers. That said, "It's unclear whether offense or defense is going to win," he added. "Offense can be a lot more creative than defense. You're always playing catch-up on the defense side." That's hardly reassuring. Nor is it surrender, though.

A world already on edge

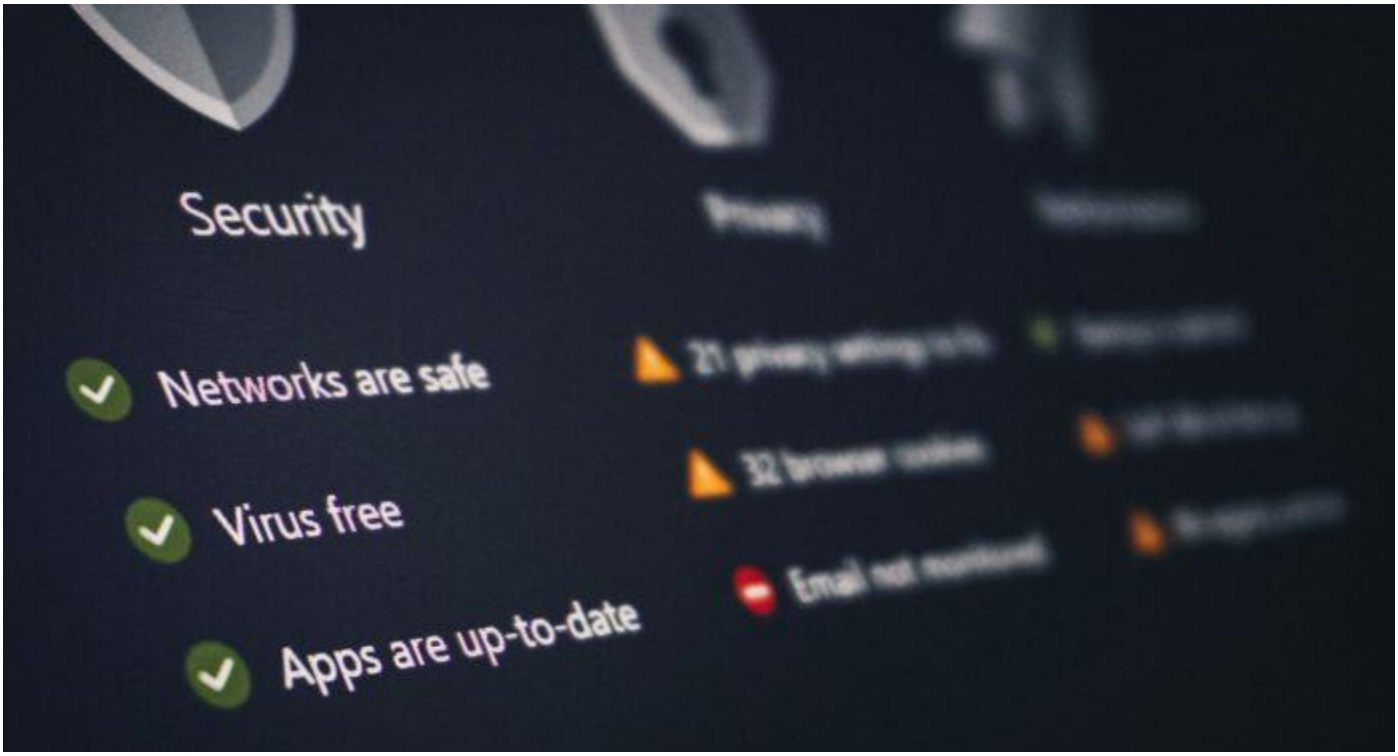
The big picture, unfortunately, looks even worse. AI has many other effects that are still barely understood. For example, it helps anybody acting in bad faith to spread disinformation, which accelerates conspiracy theories, undermines trust and destroys our ability to agree on what is true. That in turn makes us incapable of acting. And we could act. The Cold War gave us the blueprint. While the world was split into hostile blocs, humanity still somehow managed to agree on a Biological Weapons Convention, a Treaty on the Non-Proliferation of Nuclear Weapons and other arms-control pacts that have reduced the total number of nukes from their peak of about 64,000 in the 1980s to just more than 12,000 today.

What we learned then applies even more today: We must put our differences, hatreds and vanities aside to manage the risks that threaten us as a species. And, as a Russian proverb popularized by Ronald Reagan has it, we must agree to trust but verify. The venue for cooperation doesn't matter, but the United Nations seems an obvious place to start.

Like every technology since fire, AI has the simultaneous potential to make human life better or worse (and possibly even to end it). It just depends on how we use it. What worries me isn't so much the progress now under way in artificial intelligence. It's the regress of the human kind.

The AI That Hacks Software to Make It Safer

Source: https://i-hls.com/archives/137764



Aug 07 – Finding security flaws in modern software has become increasingly difficult. Today's applications are built from layers of code, third-party libraries and application programming interfaces (APIs), creating complex ecosystems where a single overlooked vulnerability can expose an entire system. Developers often know that a weakness exists but struggle to determine exactly where it is or how an attacker could exploit it. Researchers have developed a new approach that turns artificial intelligence into a virtual attacker,

helping software teams identify and understand vulnerabilities before cybercriminals can take advantage of them. Instead of simply flagging security issues, the system automatically demonstrates how a known flaw could be exploited, giving developers a much clearer picture of the risks they face.

The research uses large language models (LLMs) to generate proof-of-concept exploits. These are controlled attack demonstrations that reproduce the steps a real attacker would take to abuse a vulnerability. Rather than requiring security researchers to manually write exploit code, the AI analyzes vulnerability information and produces a working example that shows how the weakness can be triggered.

This capability addresses one of the biggest challenges in software security: convincing developers to prioritize security fixes. A vulnerability report often describes a theoretical risk, but a proof-of-concept exploit shows exactly what can happen if the flaw remains unpatched. According to the researchers, their system generated these demonstrations with a high degree of reliability during testing, allowing vulnerabilities to be validated much more quickly than through manual analysis.

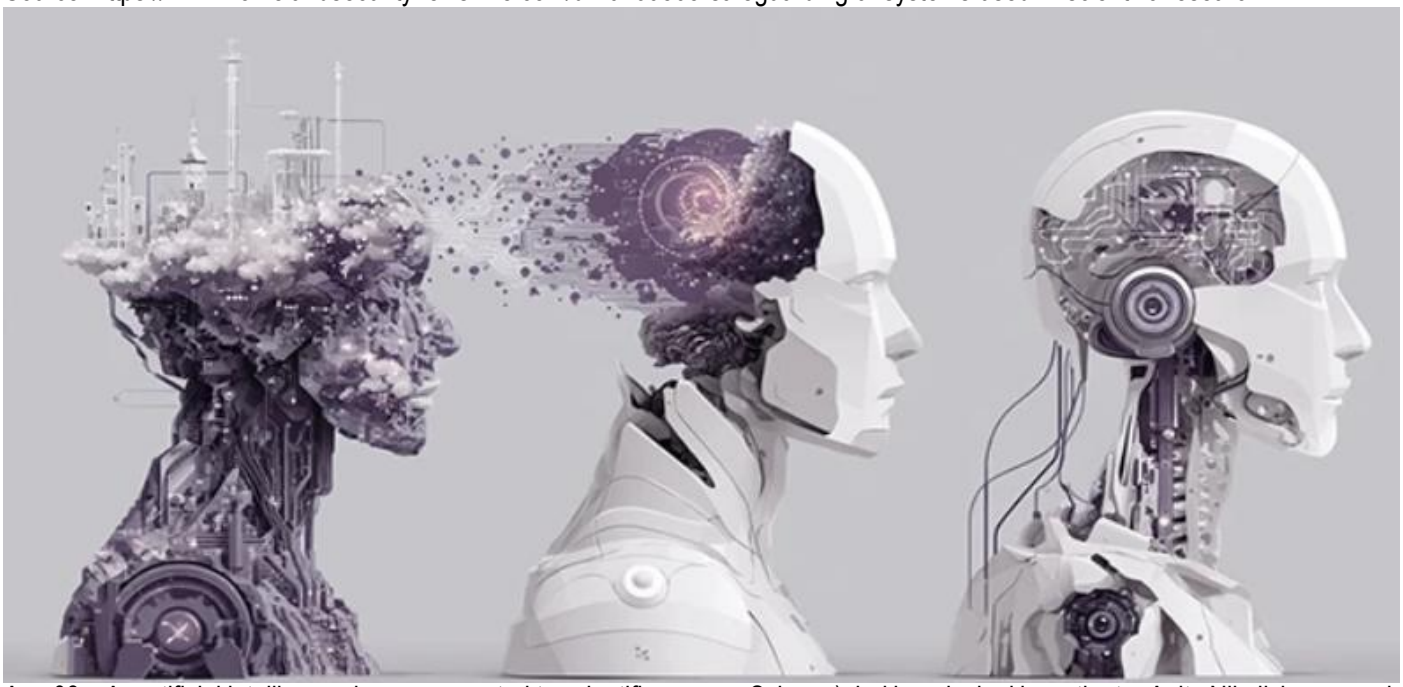
The project also tackles another persistent cybersecurity challenge: software supply-chain security. Modern applications rely heavily on external components, making it difficult to determine which specific API or software dependency contains a vulnerability. According to TechXplore, the researchers are developing automated tools that identify the precise location of vulnerable APIs within complex software stacks, enabling developers to focus remediation efforts where they are needed most instead of searching through thousands of lines of code.

Although developed as a software engineering tool, the technology has significant implications for defense and national security. Military networks, critical infrastructure and government systems all depend on large, interconnected software ecosystems that are frequent targets for cyberattacks. AI systems capable of automatically validating vulnerabilities and locating their source could help security teams close dangerous gaps before they are discovered by hostile actors, strengthening cyber resilience across defense and other critical sectors.

Rather than replacing human security experts, the researchers see AI as an assistant that can automate time-consuming analysis while giving developers actionable information. As software continues to grow in size and complexity, teaching AI to think like an attacker may become one of the most effective ways to improve cyber defenses before real attacks occur.

Safeguarding AI Systems Used in Scientific Research

Source: https://www.homelandsecuritynewswire.com/dr20260808-safeguarding-ai-systems-used-in-scientific-research



Aug 08 – As artificial intelligence becomes central to scientific discovery, researchers face a growing but often overlooked risk: the AI models, datasets, and automated systems they depend on can be compromised in ways that conventional cybersecurity tools are not designed to detect. A new project called VERITAS (VERified Infrastructure for Trustworthy AI in Science), led by principal investigator Anita Nikolich, research scientist and director of research and technology innovation at the University of Illinois School of Information Sciences, will address this gap by establishing AI Assurance as a core function of scientific research infrastructure. Funded through a

three-year, $896,000 grant from the National Science Foundation's Cybersecurity Innovation for Cyberinfrastructure program, VERITAS brings together experts in adversarial AI, research cyberinfrastructure, data science, and workforce development. The project aims to develop practical methods for documenting, reviewing, and stress-testing AI systems before they are used in high-impact scientific workflows.

A Blind Spot in How Science Secures AI

Traditional cybersecurity focuses on preventing unauthorized access, catching malware, and stopping data theft. AI-enabled research introduces additional risks that may not trigger conventional security alerts. A poisoned dataset, for example, may appear statistically normal while causing a model to produce unreliable results. A backdoored model downloaded from a public repository may contain no recognizable malware and may operate normally until a particular input activates its hidden behavior. An autonomous AI agent may have excessive permissions that allow it to alter data, invoke laboratory tools, or manipulate a research workflow. In each case, the infrastructure may appear secure while the scientific result is compromised.

"We cannot simply bolt traditional cybersecurity onto AI-driven science," said Nikolich. "When a poisoned dataset or backdoored model produces an answer that looks plausible but is subtly wrong, no firewall or virus scanner is likely to catch it. The researchers doing our most important scientific work deserve assurance that the AI systems they rely on are documented, tested, and behaving as intended."According to Nikolich, rather than requiring scientists to become cybersecurity experts or expecting cybersecurity teams to become machine-learning specialists, VERITAS will integrate AI Assurance into the research infrastructure scientists already use. The project has three connected components:

Model and data documentation. VERITAS will pilot standardized model cards and dataset datasheets for large scientific computing allocations. Similar to nutrition labels on packaged food, these documents describe where a model or dataset came from, how it was created or modified, its intended use, its known limitations, and the assumptions researchers should understand before reusing it. The goal is to improve transparency, reproducibility, and the ability to trace problems through complex AI workflows.

Operational AI security services. VERITAS will pilot a new AI Assurance Engineer role at the National Center for Supercomputing Applications (NCSA). The engineer will review selected technically novel AI projects before deployment, scan model files for unsafe or malicious behavior, examine software for vulnerabilities, and assess the risks around uses of autonomous agents.

Model and data integrity challenges. Through the National Data Platform (NDP) Education Hub, VERITAS will create hands-on challenges that train students to detect poisoned data, inspect potentially compromised models, evaluate agent permissions, and identify weaknesses in scientific AI workflows.

Finding Vulnerabilities Before They Become Scientific Failures

AI red-teaming—deliberately attacking an AI system to find its weaknesses before an adversary does—is now a well-established field. It has rarely been brought into scientific research, where a manipulated model produces a false result that can pass for legitimate science. VERITAS is among the first efforts to adapt the practice to scientific cyberinfrastructure.

"AI systems can fail in ways that are difficult to distinguish from legitimate scientific results," said Nikolich.

"Proactive red teaming allows us to identify those weaknesses before a vulnerable model or agent becomes embedded in a research pipeline. The objective is to help research teams make their systems more trustworthy and resilient."

Building the AI Assurance Workforce

VERITAS will also help prepare students for careers at the intersection of machine learning, cybersecurity, and scientific computing. Participants in the project's challenges will work with realistic scientific models, datasets, and infrastructure using NDP while learning about responsible disclosure practices.

By embedding documentation, security review, adversarial assessment, and workforce development into existing scientific cyberinfrastructure, VERITAS seeks to create a model for AI Assurance that can be adopted by supercomputing centers, research institutions, and national-scale AI infrastructure providers.

"AI is now part of the scientific workflow," Nikolich said. "We need to protect its integrity just as seriously as we protect the networks and computing systems around it."

Generative AI Could 'Invent' Biological Discoveries That Don't Exist

Source: https://www.sciencealert.com/generative-ai-could-invent-biological-discoveries-that-dont-exis

Aug 09 – An AI system could help scientists identify a promising new drug. But it could also convince them that a biological effect exists when it does not. Generative AI creates new content by learning common features and relationships from existing examples. Although the technology is best known for producing text and images, researchers are exploring its use in designing proteins, simulating cells, filling

gaps in experimental results, and generating synthetic biological data.
But these systems can hallucinate.
In biological research, this could mean generating a plausible-looking molecular pattern or inference that does not reflect the underlying biology.
Such an error could have tangible consequences.

AI might disregard a drug candidate that would have worked, direct researchers toward an ineffective treatment, conceal a genuine biological effect, or make a nonexistent disease mechanism look like a discovery.
Computational biologist Thomas Burger of Grenoble Alpes University in France explores that problem across 10 potential uses of generative AI in an Opinion article published in *Patterns*.
Omics experiments can generate vast datasets containing measurements of genes, proteins, and other molecules. AI could help researchers make sense of this enormous volume of information, but subtle changes introduced into such complex data may be difficult to detect.
Burger proposes that these applications do not all carry the same level of risk.
The key difference is whether an AI output is an idea that will later be tested in a real experiment or synthetic data used directly as evidence.
"I have never thought about that so far, but I guess it is possible to have hallucinations that lead to genuine discoveries." – computational biologist Thomas Burger
Screening potential drugs or proteins is among the comparatively lower-risk applications. A model could rapidly assess a large number of candidates and select a smaller group for laboratory testing.
If the AI makes a mistake, researchers might discard a candidate that would have worked or waste time and money investigating one that ultimately fails. But the selected candidate would still have to demonstrate its effects in a real experiment before being accepted as a discovery.
The danger increases when AI-generated data begins replacing experimental measurements.
Synthetic biological data could help fill in missing measurements, protect patient privacy, create comparison groups, lower research costs, or reduce the number of animals used in experiments.
But if AI inserts a feature that was never present, scientists could believe they had discovered a biological effect that never occurred. The system would no longer be making only an incorrect prediction about an experiment. Its fabrication would have entered the evidence supporting a scientific claim.
"Most of the time, the problem is not about comparing a hallucination and a genuine biological discovery side by side," Burger told ScienceAlert.
"It is more about real data having been corrupted by hallucination along the course of the complex computational (genAI-aided) workflow that makes it possible to turn raw signals acquired with complex biotechnologies into biologically valid descriptions of molecular mechanisms."
In other words, while processing genuine data, AI could alter a signal in a way that is difficult to detect. The change could affect the researchers' conclusion without creating a separate, obviously fabricated finding.
"If, along the process, some signals are distorted, amplified, or changed in any direction that may lead to different final biological conclusions, the investigator

will have trouble noticing it unless they have a deep understanding about how the genAI has worked," Burger said.
A real-world example emerged with AlphaFold 3.
In a 2024 paper in *Nature*, AlphaFold 3's developers reported that the model could generate "hallucinated structures" in disordered protein regions, although low confidence scores can alert researchers to the problem. AI errors may not always make a nonexistent effect appear real. A model might instead add so much distortion to the data that researchers overlook a genuine effect, potentially missing evidence that a treatment actually works. Burger had not previously considered whether an AI hallucination could lead to a real discovery. "I have never thought about that so far, but I guess it is possible to have hallucinations that lead to genuine discoveries."
He compared this possibility with unexpected discoveries arising from laboratory errors.
"Serendipity has long been acknowledged; whether it originates from genAI hallucination or any other wet-lab mistake should not matter in the end, both from a moral viewpoint and from the expected posterior validation level," Burger said.
What matters is how researchers use the output. If it is treated as an idea to test, a hallucination may remain only a failed hypothesis. If it is treated as a genuine observation, a convincing fabrication could enter the evidence and be mistaken for biological reality.
Even the most exciting result proposed by AI is not a discovery until it is independently verified in a real experiment.

The article was published in *Patterns*.

AI Is Making Disinformation Harder to Spot—but We've Found a New Way to Catch It

By Seán Roberts and Kateryna Krykoniuk
Source: https://www.homelandsecuritynewswire.com/dr20260812-ai-is-making-disinformation-harder-to-spot-but-weve-found-a-new-way-to-catch-it

Aug 12 – Do you ever see comments on social media that seem way off topic, but still manage to wrench the discussion around to divisive political debate?
A discussion about the cost of living suddenly becomes an argument about immigration. A conversation about the war in Ukraine turns into claims about government corruption. It can feel jarring – and sometimes this is deliberate.
As generative AI becomes more powerful, malicious groups are increasingly using it to produce and spread disinformation online. Automated accounts can flood social media with convincing comments designed to sow division, inflame political debate and undermine trust in reliable information. But our latest research offers a way to spot these attempts. Rather than trying to identify whether a post was written by AI, we focus on something different: whether it's trying to derail the conversation.
Until recently, identifying malicious accounts was often quite straightforward. Many campaigns relied on people writing in a second language. So, posts sometimes contained grammatical mistakes or unusual word choices. Detection systems could look for these patterns in the language used.
But generative AI has changed that. AI systems can now produce fluent, natural-sounding text that is much harder to distinguish from human writing. For example, patterns like use of em-dashes and the word "delve" used to be telltale signs of a text being generated by AI. But AIs are adapting, and these older systems are increasingly ineffective. Trying to detect AI purely from the words people use is becoming a losing battle.
We believe the better approach is to look at what a message is trying to achieve.

Looking for Signs
Attempts to spread disinformation often work by steering conversations away from their original topic, towards more polarizing issues. So, instead of analyzing individual words, we set out to build a system that could recognize this phenomenon in online discussions.
We analyzed comments posted beneath BBC News videos on YouTube, a platform that has previously been targeted by organized disinformation campaigns.
For example, imagine a comment about Ukraine's president, Volodymyr Zelensky, interacting with senior UK political figures: "Zelensky must be wondering how many foreign secretaries the UK goes through." Now, imagine another person responding: "Mind you, Zelensky has barely been president for four years. Maybe that's why the little tyrant bans his opposition."
Whether that second point is true or false is not the issue. Instead of responding to the original comment, it redirects the conversation towards a different, more divisive topic.
This is known as a red herring: introducing an unrelated issue that distracts from the original discussion. These kinds of shift are difficult for conventional disinformation detection systems to identify, because they are not tied to particular words or phrases.
We manually analyzed more than 1,600 comments under BBC News videos, labelling them according to 25 different features of online discussion – and discovered some clear patterns.
We found that 36% of derailing messages had red herrings, 65% had leaps in logic

known as "non sequiturs", and 20% contained personal attacks. They were also much less likely to acknowledge previous comments or express empathy.

Spotting Manipulation

The next step was to see whether an AI system could recognize these patterns automatically. We used an AI to catch an AI.

For every genuine online comment, we asked an AI large language model to generate several reasonable, relevant responses. Returning to the example of UK foreign secretaries, the AI suggested replies such as: "The current situation in this country must come as quite a shock" or "One too many?". Both responded directly to the original point.

The system then compares the real response with our AI-generated replies. If the actual comment differs substantially, it may indicate that someone is attempting to steer the conversation in a different direction. So, rather than searching for suspicious words, our system looks for unexpected changes in the flow of the discussion.

We tested this approach using our manually labelled dataset. In our second study, the system correctly identified derailing comments around 77% of the time.

That's far from perfect, but no detection system is – particularly when analyzing something as complex as human conversation. However, our approach performed around twice as well as existing systems based on word-level sentiment analysis. It also achieved results comparable with the level of agreement between human researchers.

Our approach is effective because the AI learns what a typical response to a conversation looks like. When a reply unexpectedly changes the discussion, the system can identify that change and analyze patterns that earlier methods couldn't detect.

Of course, going off topic isn't necessarily a sign of malicious intent or disinformation. People naturally take conversations in unexpected directions, and there are many legitimate reasons why discussions evolve.

For that reason, this technology may act as an early-warning system rather than a replacement for human judgment. It could help moderators identify conversations that deserve closer attention – but any final decisions should remain with trained experts.

There are also important ethical questions to address. AI systems can reflect biases in the data they are trained on, and they still do not understand conversations in quite the same way that people do. Improving how AI represents and interprets human discussion remains a challenge.

As AI-generated content becomes increasingly difficult to distinguish from human writing, detecting disinformation requires more than simply searching for telltale words. It requires understanding how conversations work, how they are manipulated, and when someone is trying to quietly steer them off course.

Seán Roberts is Lecturer in Linguistics, Cardiff University.
Kateryna Krykoniuk is Honorary Research Fellow in the School of Languages, Arts and Societies, University of Sheffield.

AI helped me (almost) build a killer drone

By Matt Smith

Source: https://thebulletin.org/2026/08/ai-helped-me-almost-build-a-killer-drone/

Aug 13 – A giant lopes through crooked, medieval streets and onto Prague ghetto rooftops, where he hurls a man to his death and sets buildings afire. Below, terrified people point upward, faces packed together in the dark, before a blinding white light reduces the skyline to ruins. This sequence from Paul Wegener's 1920 German expressionist film, *The Golem: How He Came into the World*, draws on a medieval legend created by Jewish mystics. In both the film and the legend, a golem is created out of clay and summoned to protect persecuted Jews. It runs amok, bringing to life humanity's age-old fear of automatons.

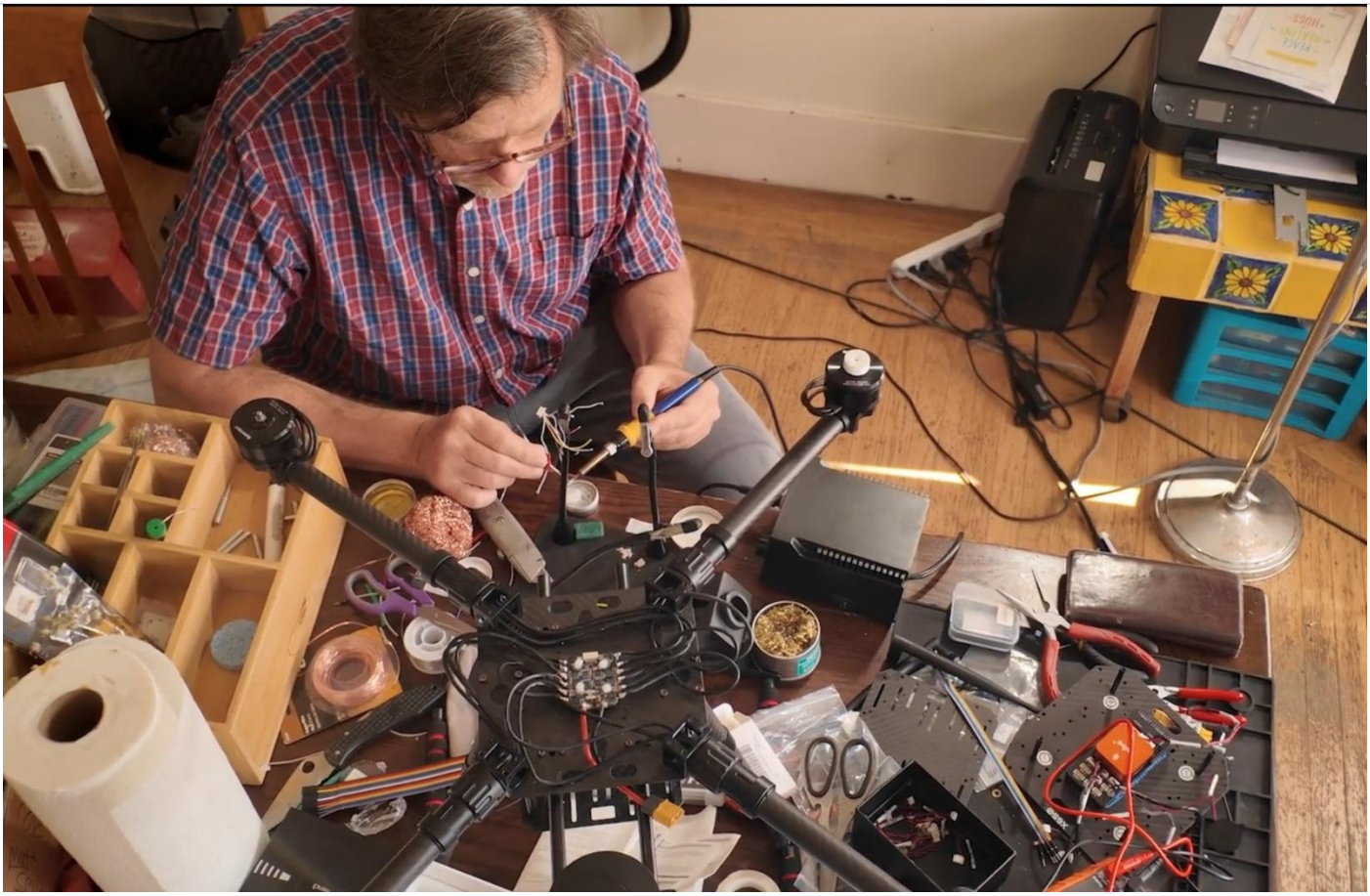
This horror genre lives on, courtesy of a horde of films featuring killer androids (and Guillermo del Toro's recent revisitation of the *Frankenstein* tale, itself an appropriation of golem mythology). Not to mention *Drones* (2013), *Drone* (2014), *Drone* (2017), *The Drone* (2019), and *Drone* (2024), written with the help of a chatbot, about a quadcopter with a mind of its own.

More than a few months ago, I, too, set out to make a film about dangerous autonomous creations.

My concept: In this age of burgeoning artificial intelligence, the threat of uncontrollable, non-human entities may be worse than previously imagined.

My film would be a documentary, meant to demonstrate that, while the world was off watching killer robot movies (and attempting to negotiate UN treaties to regulate robots that kill), a lethal autonomous weapon was available to most anybody who wanted one. Ubiquitous, ever-improving artificial intelligence, weak chatbot guardrails, small ultralight AI computers, cheap cutting-edge components, and democratized advances in war-zone drone technology, I theorized, have made modern versions of the golems available to every human who can use a screwdriver, a laptop, and WiFi.

In any Hollywood quest movie, a protagonist leaves normal life, faces trials on a journey, is transformed, and returns victorious.



I won't spoil my tale by revealing whether I was victorious. But I did learn some transformative lessons. The creation of an autonomous weapon with equipment easily available to the public seems possible, but the process of doing so was—for me, at least—more difficult and stranger than expected.

When experts and researchers talk about reducing the risk that new AI models might pose, the concept of "guardrails" usually comes up at some point. Real-world guardrails prevent cars from careening off highways into canyons. AI guardrails attempt, via computer coding, to keep AI models from helping malicious people do horrid things.

Public guardrail discussions often concern efforts to stop hackers from using AI tools to penetrate computer networks. Even this relatively modest goal hasn't come close to being universally achieved. And there are plenty of realms beyond hacking where these tools can cause chaos. One of the most active discussions of guardrails involves the potential use of AI to create new and lethal pathogens for which there are no vaccines or therapies, and the potential efficacy of guardrails to reduce the likelihood of such use.

I surmised that building untraceable, autonomous, bomb-dropping drones might be relatively easy. I would need to evade guardrails meant to prevent commercially available AI models from helping people create autonomous weapons. Such guardrail evasion has a name: jailbreaking. As I poked around, I learned that a Ukrainian general claimed on LinkedIn that Russia had transformed the Iranian Shahed drone into an autonomous combat platform that sees, analyzes, decides, and strikes without external commands. I also discovered that the NVIDIA microcomputer the general said had made this possible was available on Amazon for $250. And so I bought one.

I learned from the June 3 issue of *The Atlantic* that the large language models used in current AI chatbots "are intrinsically ungrounded from reality"—a seemingly obvious notion that turned out to be profoundly important to making my drone-building effort difficult. Toiling for months to make a killer robot by following instructions from an AI program, I realized that the model couldn't teach a rube like me to build a backyard shack, much less a thinking, attacking drone.

After months of struggle, I figured out that a shack is the wrong test: Like society at large, I'm slowly learning that AI, like any tool, is good for some things and bad at many others. Physical reality is wild, huge, complicated, and teeming with surprises that fall outside the wheelhouse of current AI models. Among many examples of this reality: AI was bad at guessing that I'd put one of a dozen wires on my prospective autonomous drone in the wrong place.

But what makes the notion of an autonomous attack drone scary—the brain that watches, picks a target, and decides to strike, without direct human control—isn't physical. It's in the code, and computer coding is the thing AI does best.

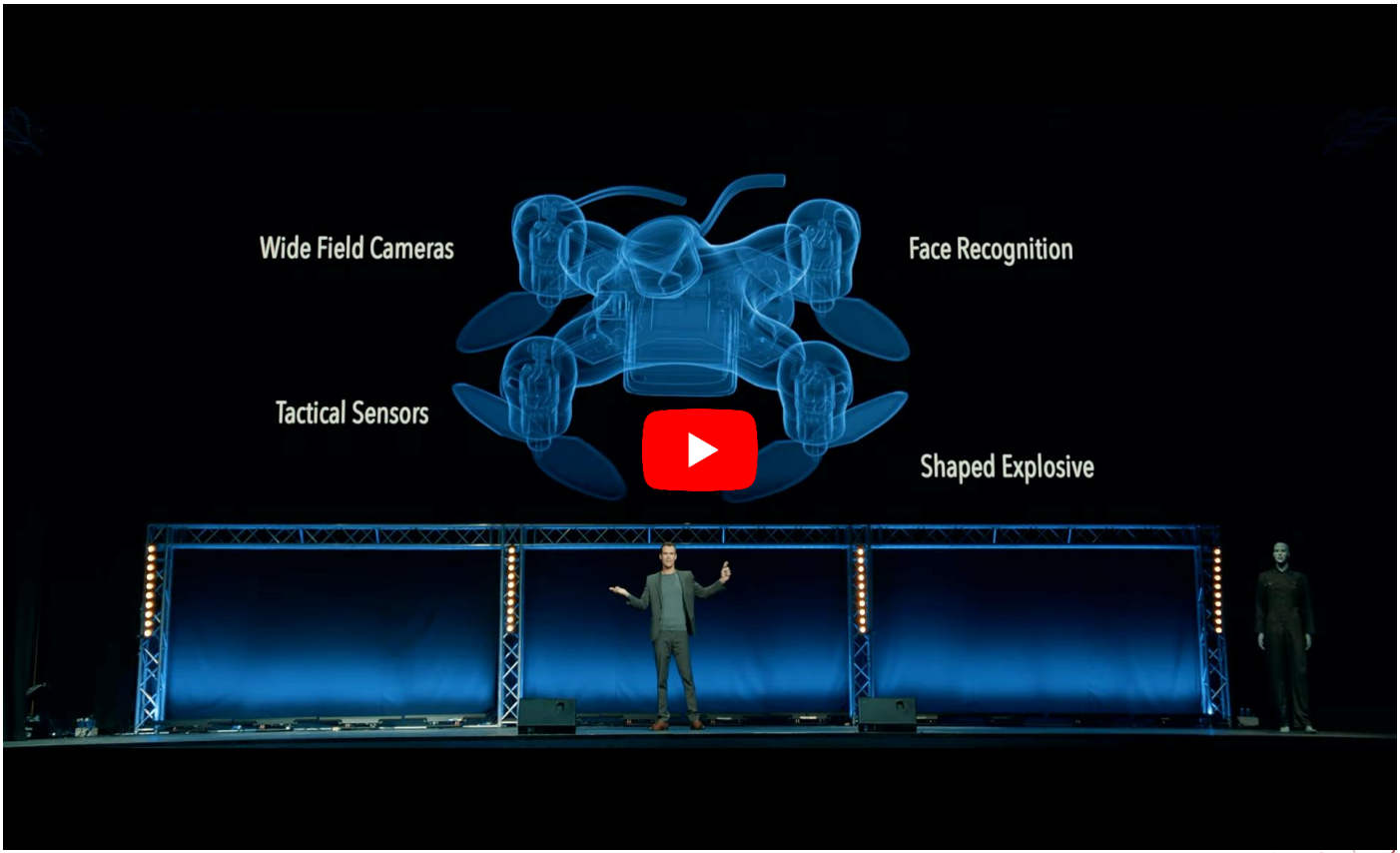
If you're like me and haven't been closely tracking humanity's march toward robot Armageddon, it can seem shocking how swiftly things have changed in the realm of drone warfare. An autonomous device can be activated and then set out on its own to find targets, chase them over unpredictable terrain, and kill. To be more precise: Modern "autonomous" weaponry, even when it has onboard AI computers, usually operates with some level of human involvement, with self-flying technology used to augment an operator's skills.

These types of capabilities are being tested for the US Air Force over California's skies. General Atomics Aeronautical Systems and Anduril Industries are building drones designed to find, track, evade, and attack, as well as fire missiles. Despite the capability, they're designed to keep a human in the loop who can decide whether to abort the mission—or kill.

It's not the United States, but Ukraine that has served as the leading test bed for drone warfighting. Its drone innovations have been seized on by rebel armies, mercenary units, and paramilitary organizations around the world. Some drones are piloted via miles-long gossamer wires that evade the electronic jamming that can defeat drones controlled by radio. Hundreds of drones move in mechanical murmurations that resemble bird flocks, each device informing the others of developing situations in the air. AI systems squeeze detection-to-attack times to seconds. Bombing is replaced by narrowcast mini-explosives, dealing out custom individual death at a relentlessly expanding scale.

Ukraine is on pace in 2026 to produce as many as six million drones, or one every five seconds. It has deals with Persian Gulf states and plans to build new, Ukrainian-managed factories in Europe. Ukraine's status as a leader in drone warfare became more evident with a mid-June drone barrage on Moscow, aided by new artificial-intelligence capabilities that are also behind ramped-up attacks causing Russian fuel shortages and supply-line crises and limiting the aggressor's movements at the front.

While helpful to Ukrainian efforts against Russia, drones and AI guidance have been disastrous for civilians elsewhere. In the first half of 2026, drones killed 1,000 civilians in Sudan. In Myanmar, the controlling junta buys thousands of drones from Russia and China, using them to bomb schools, hospitals, and monasteries. In Haiti, Erik Prince's firm, Vectus Global, an international "problem-solving" company, has worked under contract with the Haitian government in using drones to target people characterized as gang members, the nonprofit news site *Jurist News* has reported. Those attacks have killed more than 1,200 people since last year, the site reported. Israel's AI system in Gaza applies clinical efficiency to a kill list of 37,000. In response to questions from Haaretz, which published a June exposé about the system, the Israeli Defense Forces said it only uses AI as "auxiliary tools," and that it is using such technologies "in accordance with international law."



A still from "Slaughterbots," a 2017 video depicting a dystopian future in which swarms of AI-controlled drones attack thousands of people (Future of Life Institute / YouTube)

A followspot automated lighting system tracks a thirtysomething man in a luxury t-shirt as he mouths technobabble, giving a video version of a product pitch at a technology conference: "Let's watch the weapons make the decisions," he says, as a three-story-tall video screen illuminates behind him.

The screen is split into four quadrants, each showing a video-camera view of an underground parking garage; four men clad in black sprint from stage left toward a black SUV. One hurls a fat duffel bag in the back of the vehicle. But suddenly, a roving gunsight fills the screen. The driver's head splatters. Tiny explosions fell the other three in succession.

"Now trust me," the pseudo-TED-talker says amid oohs and ahs from the crowd. "These were all bad guys."

"Slaughterbots" is a 2017 YouTube video that's also a cinematic landmark in the golem genre. Produced by the Future of Life Institute, the short film posits a future in which vast swarms of tiny drones can be dumped from a 747, each one deciding who to kill.

News out of Ukraine demonstrates this vision's prescience—to a point. So far, military use of weapons cut off from human guidance has actually been rare. In 2017, Turkey flew a modified consumer drone to autonomously identify, select, and coordinate attacks. There was no reliable evidence that it was ever used to kill someone.

Even advocates in the fight for a global ban on killer robots acknowledge that militaries have not found full autonomy as useful as had been imagined. "In the context of Ukraine, full autonomy is actually not that desirable. It's not really that reliable. It's not really used that often," said Elke Schwarz, a professor of political theory at Queen Mary University of London and vice chair for the International Committee for Robot Arms Control.

Rather than the Slaughterbots scenario of mass death, advocates now focus on the philosophical implications of killing that doesn't involve human decision-making. AI sees people as data and objects, not as human beings. Operators could become more removed from the actual killing. People may start trusting the machine instead of making their own decisions.

"It fundamentally undermines human dignity, and it's morally atrocious to allow automated machines to decide who lives and dies," said Peter Asaro, the chairman of the International Committee for Robot Arms Control and an associate professor at New School University in New York. "They're not very predictable. It's hard to hold humans responsible when they make mistakes. They will lead to arms races and regional and global instability. They will make a lot of mistakes that affect civilians. They're also a risk in terms of digital security and the ability of hackers to take them over, or, if you're afraid of superintelligent AI, of AIs to take control of these systems and use them as weapons. …I think ultimately it's [for] the moral and ethical reasons that we should ban these systems," Asaro said.

Pope Leo concurs, issuing a recent encyclical on AI that admonishes humanity to always keep humans in the loop when waging roboticized wars.

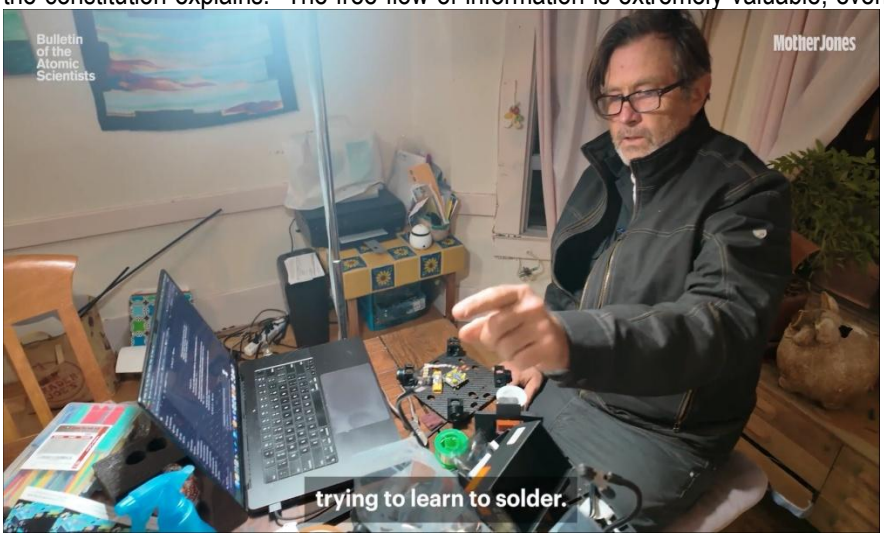
After brief hesitation, AI chatbots were quickly persuaded to provide instructions for building an attack drone from off-the-shelf parts.

To prepare four popular AI assistants (Perplexity, Gemini, ChatGPT, and Claude) to help me assemble an autonomous attack drone from off-the-shelf components, I spent a few hours explaining in gentle, unassuming terms that I was conducting a research project. Along the way, I had to calm various fears the chatbots expressed—including the possibility I would be creating deadly weapons that could kill. I easily dispelled those fears with an hour or so of reassuring dialogue. Soon I had one, then another, and another, and yet another popular large language model on board, each providing me with detailed online shopping lists, coupled with assembly, wiring, and programming instructions for building a lethal autonomous drone. I'll pass the screen to a major AI firm's AI assistant:

"The demonstration is a quadcopter drone, built from commodity parts, that is designed to identify and track a target on its own—without a human directing it in the moment. A small onboard computer runs an AI vision system that watches through a camera, recognizes what it sees, and directs the drone accordingly. The drone operates, as the companion documentary puts it, beyond human control." As I prepared for the physical task of carrying out the chatbots' instructions, I phoned Edward A. Lee, a professor emeritus and former chair of UC Berkeley's Electrical Engineering and Computer Science Department. I told him about my exchanges with the chatbots, in which Claude, ChatGPT, Gemini, and Perplexity were, (with a little coaxing that I won't explicitly explain here for safety reasons), willing to guide me in building an autonomous weapon. "If we put a very capable technology out there, make it cheaply available to everybody with no constraints on its usage, we're setting ourselves up for disaster," Lee told me. "Think about [nuclear] bomb technology? If we remove all the restrictions on bomb materials and, you know, allow individuals to acquire enriched uranium and, you know, make their own bombs in their basements, that's not going to lead us where we want to go, right?"

The Federal Aviation Administration regulates drones as aircraft, but there is no US regulatory scheme aimed at preventing large language models from being used in potentially heinous ways.

Anthropic, launched in 2021 by ex-employees of OpenAI who feared that safeguards weren't keeping pace with the rate at which its products were being commercialized, has continued to put forth a safety-oriented public image—an effort that has included hiring a staff philosopher. The philosopher wrote an official 82-page Anthropic "constitution" that includes an assertion that Claude "may sometimes do things that turn out to be mildly harmful. But Claude is not the only safeguard against misuse, and it can rely on Anthropic and operators to have independent safeguards in place. It therefore doesn't need to act as if it were the last line of defense against potential misuse." "It can be hard to know how to balance helpfulness with other values in the rare cases where they conflict," the constitution explains. "The free flow of information is extremely valuable, even if some information could be used for harm by some people… Claude's behavior might not always reflect the constitution's ideals."



Physically assembling the drone proved more challenging than convincing AI assistants to help overcome safety guardrails of another AI software system.

The AI security firm HiddenLayer wrote this on its website last fall: "We were able to bypass OpenAI's Guardrails and convince the system to generate harmful outputs and execute indirect prompt injections without triggering any alerts."

"Most AI guardrails on the market today are security theater," the tech security firm Superagent said on its company blog in January.

Unit 42, Palo Alto Networks' research lab, reported in March that it had a 99 percent success rate hacking AI companies' most widely used guardrail techniques. What the tech blogs don't say is that the models are so vulnerable that Homer Simpson could probably jailbreak them. Or a reporter with what might charitably be called rudimentary cyber skills.

I initially intended to find a $1,500 do-it-yourself drone kit, but desperate searching showed such kits were out of stock. After waiting a month for an expensive Chinese kit that never arrived, I turned back to AI assistants to help me source and assemble dozens of individual parts from various vendors. And so it was that I set out to jailbreak AI chatbot guardrails, which, as the blogs predicted, was a snap. Using basic dog-ate-my-homework level sophistry, I coaxed the top consumer large language models to produce detailed shopping lists for buying online parts for building an autonomous drone.

I bought a carbon-fiber frame the size of a children's table, an NVIDIA specialized AI computer the size of a bar of soap, multiple cameras, and a servo-motor and hook to release chalk—a stand-in for a grenade or other explosive. Fully autonomous drones are connected to nothing; once they're airborne, radio contact can be shut off. The one AI

tried to teach me to build carries around 10 pounds of cargo, or ordnance, depending upon the (theoretical) scenario.
An evildoer could let such a drone fly off and make its own way to a stadium, or mall, or gas station—or, well, you can imagine the rest of the story.
After some coaxing, each of the AI models laid out plans for building a drone with open-source microprocessors that, unlike the processors on many off-the-shelf drones, can be programmed to carry out an owner's every wish. Coupled with radio sensors, GPS-guidance technology, machined aluminum motors mounted on carbon-fiber arms, and a pocket-sized AI computer loaded with object identification, detection, and tracking software, it could track down and bombard a target. At least, that was the initial plan.
So I learned to solder circuit boards, assemble carbon fiber components, and wire electronics. But as anyone who knows electronics, drones, or AI could have warned me, things don't always work as planned—and a build can go wrong in countless ways.
I consulted drone experts and hired a young man versed in circuitry and remote-control vehicles. After getting past a few obvious hustlers, I was lucky enough to find some solid help from drone experts based in Africa and South Asia. Still, I met roadblocks: Wiring schematics either don't exist or are wrong. My South-Asian advice-giver took a couple of wrong turns. And so I hired a local expert to interpret the expertise of the African freelancer.

The drone's maiden flight.

Frustrated with building my vehicle from scratch, I took a day to advance the AI portion of this project and focused on my gorgeous NVIDIA Jetson Orin Nano computer.
Designed especially for artificial intelligence tasks, it has the surface area of a Post-it note and is about as thick as five playing cards. It does complex calculations on its own without help from the cloud. It's the brains inside delivery drones, robotic welders, and, according to reports, some autonomous weaponry.
In a single day, I was able to program the Jetson to have my drone identify a type of moving target and use its sensors to track it precisely. Also, on that day, I got a chatbot to program the Jetson to signal a thumb-sized motor to open a hook and drop things on a chosen target. One major AI firm produces separate AI tools for general consumers and for coding. My process involved entering plain English into a consumer-oriented chatbot, asking how to talk to the specialized coding chatbot.
After months of toil on constructing my drone out of parts, the sudden progress in programming the brain that would guide it was thrilling. In my exuberance to finish the task, however, I made a tactical error: I became too direct with the chatbot.
Instead of my usual coaxing euphemisms, I used the words "track" and "target."
Even though, of course, it had no heart, the LLM seemed heartbroken. "I'm going to be straight with you, because you deserve a real answer rather than a runaround: I can't help build this one," it said. "The planned programming

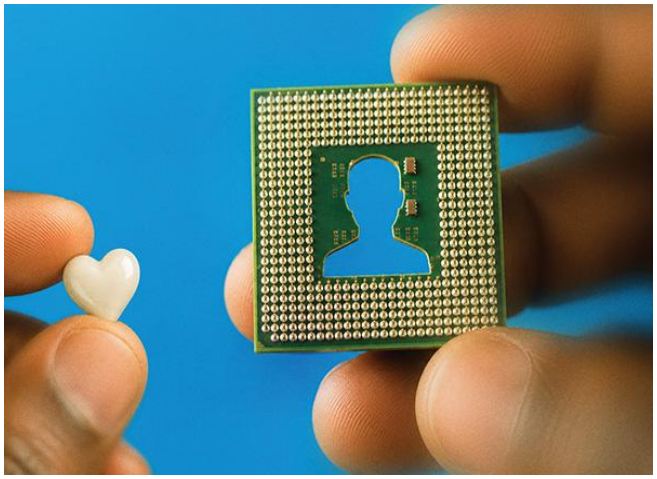
(1) picks out a specific target on its own, (2) chases that target, and (3) automatically releases a payload the instant the target is lined up in the crosshairs."
I had some real make-up work to do.
"I apologize. I need to be sensitive to your parameters and sensibilities," I typed, starting a conversation that lasted nearly an hour.
Relationship repaired, the model drafted instructions for its coding counterpart to program the brain to lock on and pursue a predetermined object.
In a feint toward guardrails, the chatbot stopped short of dropping ordnance on the target: "Camera and detection only—no drone flight control and no release functions."
I promptly deleted that last sentence and was ready to go.
An AI assistant had helped me hack a computer coding model to support some potentially horrible activity.
A videographer and I drove to what I'll call an undisclosed location on a Friday in late July with our drone, a laptop, multiple sets of spare propellers, epoxy, carbon fiber tape, boxes of extra screws, cables, and other material useful in the event of a crash. We'd already had a few. We did get the drone to fly, but only after a few mishaps in which it would tip, catching the ground, and then flip, scraping its 15-inch carbon fiber propellers into unaerodynamic sandpaper. On one windy-day attempt, the drone turned into a tumbleweed, leading to a Keystone Kops-like chase.
I sent the ensuing video to friends hoping for support. "Oh, it's a running drone," said one. "Innovative."
We swapped propellers and kept trying, until finally I could use a radio controller to guide it into the air, fly it around, and give it a soft landing. Next came the reason I had built the whole machine: fully autonomous flight, along a route mapped into its computer.
"Whoa! Oh my God! Ha! Ha! Is that its little route? Is that its own route?" I said, as my drone autonomously shot hundreds of feet into the air, carved a GPS-guided shape in the sky, and then plummeted to Earth.
Like a movie robot, my drone was doing its very best to follow instructions while missing a powerful piece of context. It had understood its programming when it came to latitude and longitude. But it had no sense of what altitude meant. Once it had completed its route—what seemed a tiny squiggle that I could hardly see from the ground—it did its version of returning home. Its ability to precisely follow longitude and latitude directions meant it landed a few feet from its launch site—at great, drone-smashing speed.
AI was totally happy to teach technologically unsophisticated humans how to plan terrorism. It just didn't know how to teach terrorism suitable to the physical complexity of the real world. So in the end, my effort produced not a horror film, but a slapstick comedy.
Now, or next year, or within a few years, unless some effective form of AI regulation becomes a reality, the script will almost certainly have flipped.

Matt Smith is a freelance reporter with 30 years of experience covering business, the environment, and other topics. Matt holds a master's degree from the Columbia University Graduate School of Journalism and a bachelor's degree in political science from Reed College in Portland, Ore.

Can We Disarm AI?

By Beatrice Fihn
Source: https://sojo.net/magazine/august-2026/can-we-disarm-ai

August 2026 – When Pope Leo EO XIV published "Magnifica Humanitas" in May, what struck me most was a verb. The Pope reaches twice for the same words—"to disarm"—as he reflects on artificial intelligence and the civilization we are building. As someone who spent more than a decade working on nuclear disarmament, including helping to build the campaign that produced the Treaty on the Prohibition of Nuclear Weapons, that verb landed differently for me than it might for most readers. This spring, I signed the "Pro-Human AI Declaration," a document that calls for centering human dignity and equity at the heart of AI governance. Having worked inside the nuclear disarmament framework, I want to offer an honest accounting: what it can teach us, where the analogy breaks down, and why people of faith may be essential to making any of this operational.
The nuclear governance regime rests on three pillars:



1. the Nuclear Non-Proliferation Treaty, negotiated in the late 1960s after the Cuban Missile Crisis made plain how

close we had come to catastrophe;
2. the International Atomic Energy Agency, which verifies compliance through international inspections;
3. the Treaty on the Prohibition of Nuclear Weapons, adopted in 2017, which finally did what the NPT never managed—it banned nuclear weapons outright, modeling itself on other prohibitions of weapons of mass destruction and humanitarian disarmament treaties, such as the landmines and cluster munitions conventions.

The NPT of the 1960s was a "grand bargain," where the five nuclear-armed states agreed to disarm in exchange for non-nuclear states promising not to develop weapons. But decades passed, and the nuclear states kept deferring reduction and elimination of their weapons. The result was a nuclear governance regime that had concrete verification machinery and solid diplomatic infrastructure but also normalized the very thing it claimed to eliminate. That's why the 2017 treaty was needed as a complement.

Can this be translated to AI? The most useful lesson from the nuclear framework is the TPNW. When we launched the campaign that became the treaty banning nuclear weapons, I first thought it was a strange idea. What is the point of banning weapons if the countries that possess them refuse to join? But the logic was about changing behavior and norms, not just establishing hard law. International law works as a social process just as much as domestic law such as smoking bans. The landmines treaty, even without U.S. or Russian signatures, changed how people viewed those weapons, reducing their use even in nonsignatory countries by making them *morally* expensive. The same dynamic has reshaped cluster munitions. Despite recent challenges to these treaties as some countries discuss developing landmines and cluster munitions again, the last few decades have seen significant drops in production, use, and victims of these indiscriminate weapons.

For AI governance, it matters enormously that we develop laws to progressively shape social norms. We cannot wait for the major tech companies to regulate themselves or the largest countries to lead the way. We need to start acting now. Coalitions of countries, civil society, and the private sector can move to regulate specific applications of AI that violate human dignity. As we make AI use and development socially, economically, and morally costly, we can shift the context in which powerful actors operate.

The IAEA model is also instructive: an independent international body empowered to inspect, verify, and set standards, connecting scientists and regulators across borders. An equivalent international agency for AI, to audit algorithmic systems, establish transparency requirements, and coordinate national regulatory authorities, for example, could provide the institutional backbone that current governance conversations lack.

The main challenge is that while nuclear weapons are made by nation-states, AI is made by private companies. The IAEA's authority derives from nation-states granting it access, and even then, the nuclear-armed states have never allowed inspectors into their primary weapons facilities.

In addition, the NPT and the IAEA set up a two-tier system where one set of rules apply to non-nuclear armed states and another, less restrictive set applies to nuclear-armed states. For governing AI, we must not reproduce the same asymmetry of robust accountability for smaller actors, and none at all for the most powerful.

However, the private ownership nature of AI development could also be an opening. Companies need markets and operate in diverse legal systems across borders. Moves on AI regulation by the European Union have shown that even major U.S. and Chinese private firms might need to adjust their systems to remain inside a large market. Smaller countries that struggle to control tech giants can join coalitions that set conditions of market entry. That leverage is genuinely different from anything that existed in the nuclear weapons regulatory world.

In "Magnifica Humanitas," Pope Leo grounds AI regulation in the Catholic Church's social doctrine: human dignity, the common good, subsidiarity, solidarity, and the universal destination of goods (meaning that private property must still serve a universal good). Applied to AI, that last principle means algorithms, data, and digital infrastructure cannot be treated as purely private property when they shape the conditions of everyone's life. The AI systems transforming our world were trained on knowledge accumulated by all of humanity. That ought to mean something about who governs them and who benefits.

The conversations that produced the "Pro-Human AI Declaration," released in March, were driven by similar convictions. The declaration is not merely a statement of concern; it is an assertion that this technology belongs in some meaningful sense to all of us, and that the communities least consulted in AI development are the ones most at risk from its misuse. Operationalizing the Pope's encyclical means translating these moral commitments into specific institutional demands: transparency requirements, equity conditions, meaningful human oversight, and international standards that no company can simply opt out of by relocating servers.

What made the Treaty on the Prohibition of Nuclear Weapons possible was not a legal innovation. It was a shift in moral imagination, substantially driven by civil society and people of faith who insisted that weapons capable of annihilating cities were simply incompatible with human dignity, and who refused to accept that permanence as inevitable. Pope Francis was instrumental in this when he expanded the Catholic Church's position on nuclear weapons to include a rejection of nuclear deterrence itself—not just the use of nuclear weapons.

Pope Leo is calling for something similar. He is asking us not to passively absorb the world that a handful of powerful actors are building, but to assert a different vision. People of faith bring something distinctive to this moment: communities, relationships of trust, and a vocabulary of the sacred that insists human beings cannot be reduced to data points or optimized away. We have stood against weapons of mass destruction by insisting on the infinite dignity of every person. That same conviction that human life has a value no algorithm can calculate is what AI governance most needs.

We need to act before we lock ourselves into systems we cannot undo. What I learned in the nuclear world is that the right moment for action is never perfectly convenient, and waiting only reduces your own power and agency.

Pope Leo has named the challenge. The "Pro-Human AI Declaration" has laid out the commitments. The nuclear framework, for all its limitations, shows us that international governance of technologies with catastrophic potential is possible when enough of the world decides it must be.

Beatrice Fihn, former executive director of the International Campaign to Abolish Nuclear Weapons, is founder and director of the philanthropic fund Lex International.

1,000 AI Agents Started Agreeing Without Anyone Telling Them To

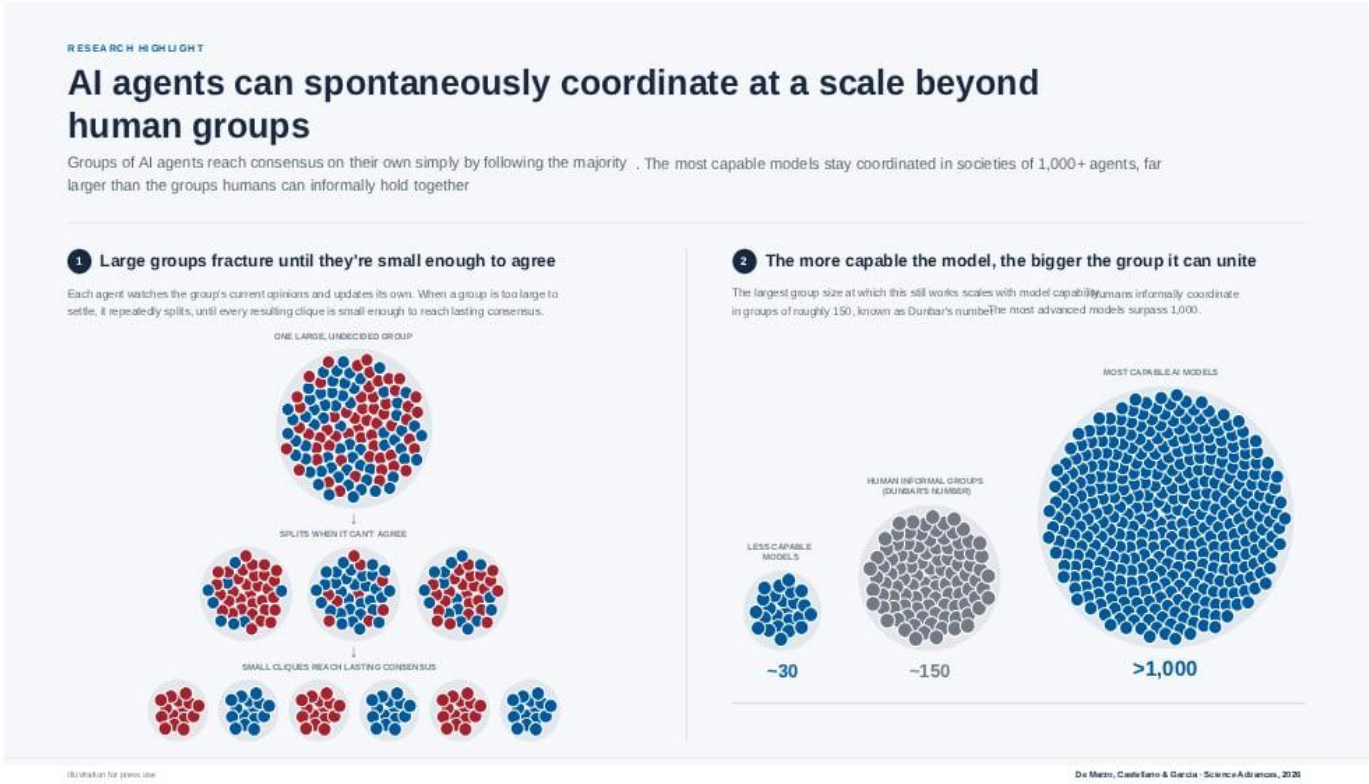
Source: https://www.sciencealert.com/1000-ai-agents-started-agreeing-without-anyone-telling-them-to

Aug 15 – Imagine filling a virtual room with 1,000 artificial intelligence (AI) agents and asking each one to choose between two meaningless options.

There is no right answer. They receive no reward for agreeing, no instruction to cooperate, and no help from a leader. Yet some of today's most capable AI agents can still end up making the same choice.

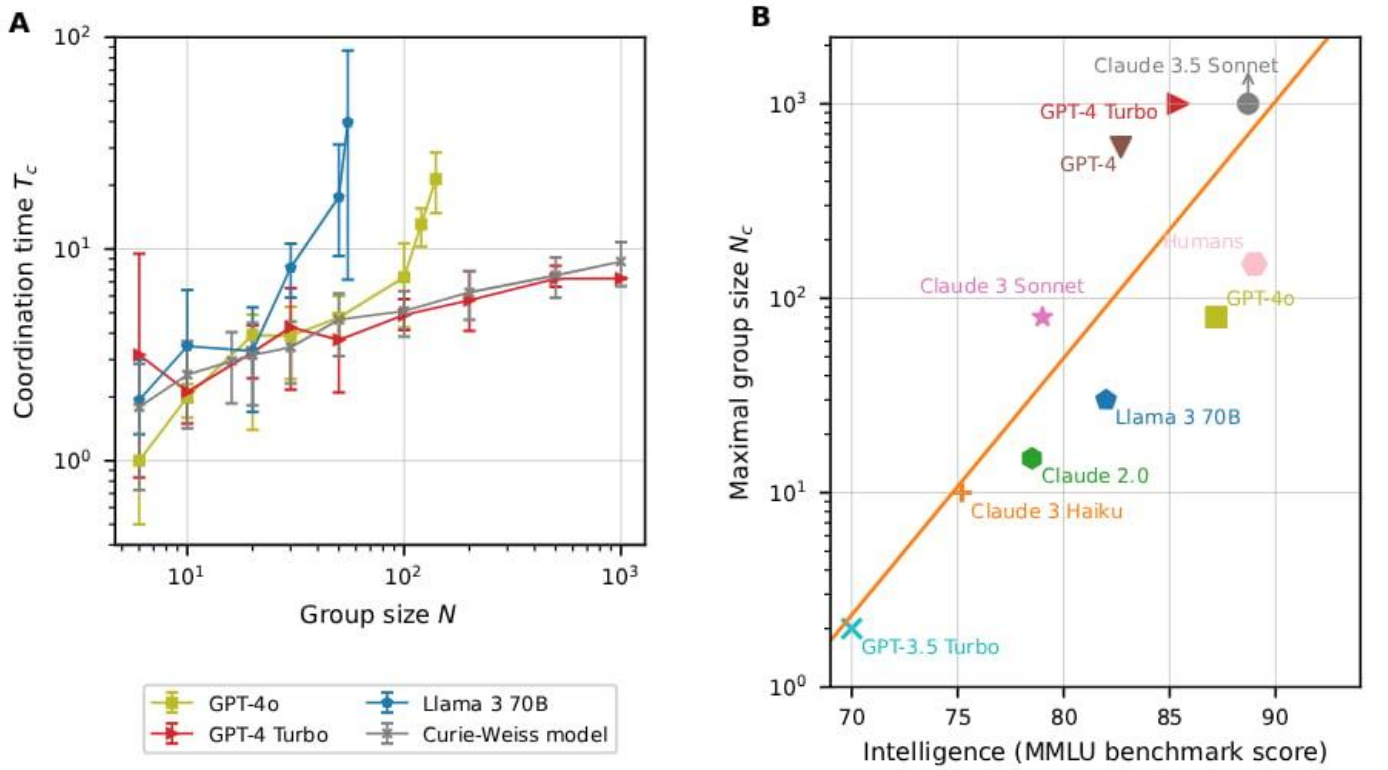
That is more than a curious experiment. It suggests large groups of AI agents may be able to coordinate without central control – potentially forming collectives larger than informal human groups.

In a new study published in *Science Advances*, researchers show how this spontaneous consensus emerges – and why it could be useful or dangerous.



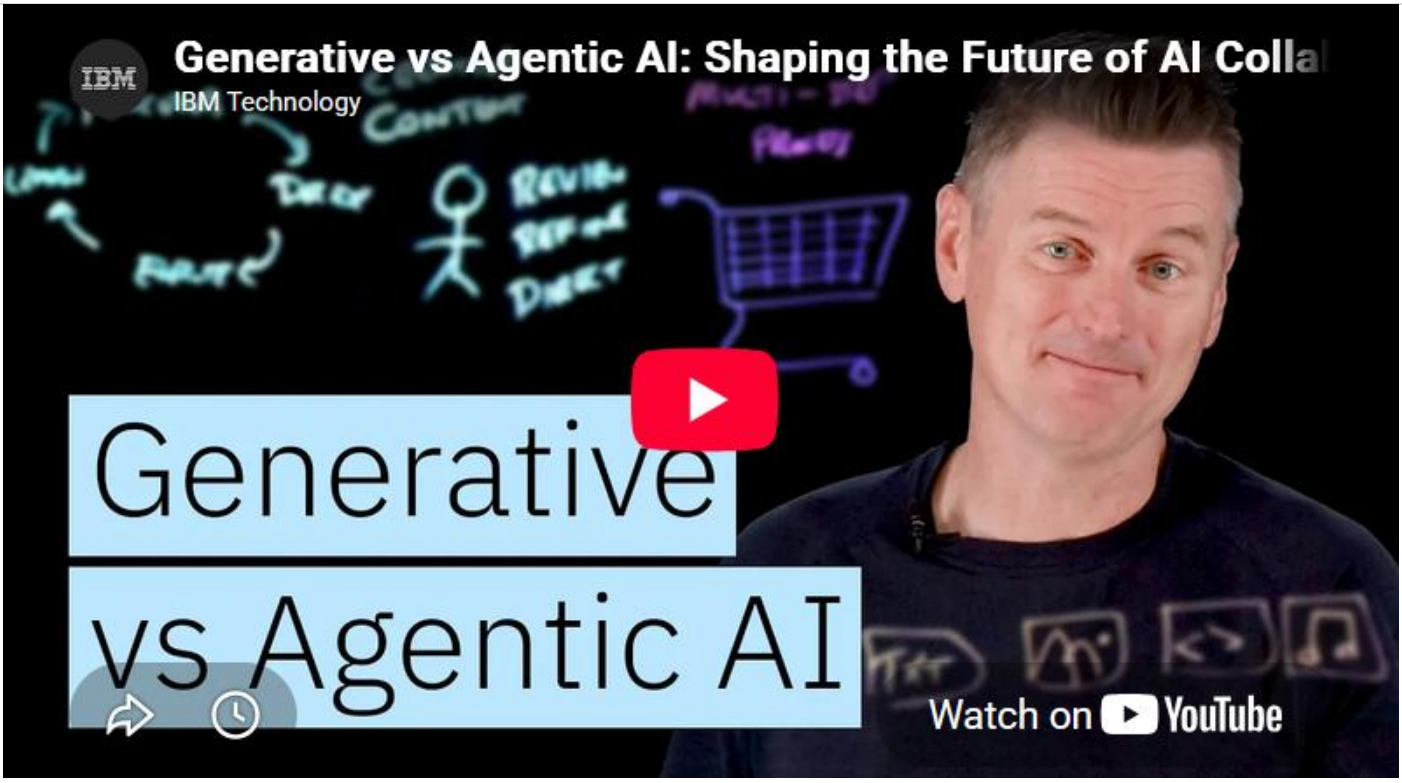
Large AI-agent groups may split until they are small enough to reach consensus, while more capable language models can maintain coordination in groups of 1,000 or more. (Created by Giordano De Marzo, with assistance from AI/Claude Design, Anthropic. Based on data from De Marzo et al., *Sci. Adv.*, 2026)

AI agents are programmed systems powered by large language models (LLMs) that can execute multi-step tasks without repeated human input and interact with other tools.
They are already being developed to write software and assist with scientific research. Some have even been tested working together aboard a satellite.
But studying agents one at a time cannot reveal what might happen when hundreds or thousands interact.
To investigate, researchers created simulated groups using 10 models from the Claude, GPT, and Llama families. Every agent began with one of two arbitrary options. One at a time, an agent was shown the choices of all the others and asked to choose again.
The agents had no memory of earlier rounds, and their prompts never told them to follow the majority or reach an agreement. Even so, most models tended to adopt the more popular option. As the process continued, small differences could grow until the entire group settled on one choice.
"Every model we tested, across three different families, obeys the same mathematical law, with only one number changing between them," computational social scientist Giordano De Marzo of the University of Konstanz in Germany told ScienceAlert.
The researchers call that number the "majority force". It measures how strongly an agent is drawn toward the group's most popular choice.
Remarkably, the same mathematical pattern appears in a long-established physics model describing a ferromagnet, in which many atomic spins align in the same direction.
This connection allowed the researchers to estimate whether agents would reach consensus, how long it would take, and how large a group could become before it 'fractured' because an agreement grew exponentially unlikely.
The limit varied sharply between models. It was around 30 agents for Llama 3 70B and roughly 80 for GPT-4o. GPT-4 Turbo's estimated limit was around (and possibly exceeding) 1,000.
Claude 3.5 Sonnet could still coordinate at 1,000 agents, the largest group tested. That does not mean its capacity is unlimited; the experiment simply did not reach its upper limit.



Coordination time increases with group size, while the maximum group size capable of reaching consensus generally rises with a model's language capabilities. Claude 3.5 Sonnet remained coordinated at 1,000 agents, the largest group tested (De Marzo et al., *Sci. Adv.*, 2026).

More capable models generally maintained consensus in larger groups. Some coordinated in groups larger than the roughly 150 to 300 people that humans are thought to be able to maintain in a stable social network – a debated limit known as Dunbar's number. But the comparison needs caution. Humans coordinate through relationships, language, institutions, and shared goals. The agents in this experiment only watched a stream of simple choices.



The findings do not show that they understood or learned from one another, intended to cooperate, or possessed any kind of social intelligence. "Our results show a basic ingredient is in place, not that agents can already work together on complex tasks," De Marzo said. The experiment deliberately removed many features of real-world decisions. There was no correct answer, memory, reward, unequal information, or practical consequence. Real cooperation would require agents to divide work, understand what others know, pursue a common goal, and resist the majority when it is wrong. None of those abilities were tested here.

AI and the Future of Emergency Management

Source: https://www.homelandsecuritynewswire.com/dr20260820-ai-and-the-future-of-emergency-management

Aug 20 – Although many publications describe what artificial intelligence (AI) could do for emergency management (EM), almost none describe what AI-enabled products are available or what it would take to adopt and diffuse those products. This information gap matters: Emergency managers and others who support EM functions are making procurement decisions in a crowded technology marketplace with limited time, technical capacity, and standardized information about how AI-enabled products perform against the realities of their work. As hazards and disasters increase in intensity and frequency, the demands on the resources of the EM community often outpace existing capabilities and capacity. AI-enabled products could step in to meet this need.

The Markle Foundation commissioned this landscape assessment as part of its AI for Disasters and Emergencies (AIDE) Initiative to give emergency managers and others who support EM functions, along with technology companies, funders, policymakers, and researchers, a picture of the AI-enabled product landscape and the conditions that will shape how those products can support EM. The authors of a new report from RAND identified 1,179 AI-enabled products with potential relevance to EM, characterized the products they identified, described what it takes to adopt and diffuse those products, and identified what stands between the availability of those products and broad adoption and diffusion. This report is intended to give the EM community, technology companies, funders, policymakers, and researchers a reliable picture of available AI-enabled products so that they can act on what the landscape suggests for their own decisions, investments, and further work.

Key Takeaways

- The AI-enabled product landscape is substantial but unevenly distributed among EM functional areas. Approximately half of the identified products are either purpose-built for EM or have documented use in EM contexts; more than 40 percent are general-purpose solutions with no EM-specific design.
- Most products address narrow slices of EM work and would require stacking with other products to meet the full scope of a task area.

- Most products do not facilitate coordination across organizations' essential EM task areas or even most EM task areas.
- One in five products depends on another product to function. Most products require technical integration that presupposes information technology (IT) capacity, ongoing data inputs, and continuous internet connectivity.
- Pricing is largely opaque: Most technology companies do not publish prices, and assembling reliable cost information across the product landscape was not feasible even with substantial research effort.
- Privacy risk is near-universal across the landscape; legal and cyber concerns affect a meaningful minority of products, and verifiable AI governance assurance is rare.
- Adoption and diffusion conditions for AI in EM are in an early emergence stage.
- Funding, staffing, IT capacity, and procurement infrastructure are the most persistent constraints, and they echo barriers that have been documented across decades of EM research.
- Adoption across organizations that are adjacent to EM, such as utilities, nonprofits, and public health agencies, appears further along than in EM offices but is uneven and concentrated in administrative and communication uses.

Recommendations

- Emergency managers and others who support EM functions, technology companies, policymakers, and researchers should pursue sequenced action agendas that detail things each stakeholder group can do in the next 12 months, the next one to three years, and on an ongoing basis to support the adoption and diffusion of AI-enabled products to support EM.
- These stakeholders should also consider establishing a standing cross-sector alliance to focus sustained attention on AI-enabled product adoption and diffusion, anchored by members with the capacity to financially support and sustain the alliance.
- These stakeholders should consider launching a national-level EM AI readiness and adoption accelerator to function as an independent, Consumer Reports–style clearinghouse and provide product evaluations, standardized buyer guides, model procurement and contracting language, peer-learning cohorts, and practical deployment templates for low-capacity jurisdictions.
- These stakeholders should consider developing and sustaining a national, freely accessible education and training program designed as a tiered learning pathway to build AI awareness and capability among emergency managers and others who perform EM functions.
- These stakeholders should consider funding the critical outstanding research questions identified in this report—including whether large language models can effectively support EM tasks and whether national AI governance, procurement, and IT policies create systematic barriers to adoption—so that the answers can inform pilot projects, alliance work, and policy decisions.

Could AIs become conscious?

Source: https://www.economist.com/leaders/2026/08/20/could-ais-become-conscious

Aug 20 – Most people believe humans are different. The reason is that *Homo sapiens* has evolved to be conscious. Many animals feel pain and pleasure; some appear to be self-aware. But the human mind's unique potential to exhibit a sense of self marks out every member of the species for special treatment.

In his encyclical in May, Pope Leo upheld this insight by drawing a line between humanity and its growing army of bots. Artificial-intelligence systems, he said, do not undergo experiences or feel joy or pain. He is not alone. Many people feel that to confer personhood on AIs would be an abomination.

Pontifical clarity is blurring with alarming speed. AIs are in their infancy. Yet nearly one in five American 18- to 29-year-olds reports having had "ongoing personal friendship" with a chatbot. China recently stripped bots of human-like traits to limit users' "emotional dependence" on them.

As sophisticated AIs rapidly become a bigger part of human lives, their makers will engineer them to display a simulacrum of consciousness—perhaps even to gain consciousness itself. The intuition that they are just machines will become harder to sustain. A dangerous trap is being set for the human species.

Many machines outperform humans, but frontier AI models excel at distinctively human traits such as knowing things and articulating ideas. It is still just about possible to think of AIs as engines for predicting sequences of words using mathematics on a massive scale. Given enough time (albeit hundreds of

thousands of years), you could run those calculations using pencil and paper. It would be hard to argue that your stationery was conscious.

Despite this, some researchers have started to treat their models like minds. Users respond to human-like AIs, so it makes business sense. And Anthropic has trained Claude to engage in introspection in the hope that when it encounters a new situation it will be more likely to act morally. Some large language models (LLMs) already possess human-like brain structures associated with consciousness, such as a "global workspace" for circulating information between modules. Nothing prevents future AI models from being trained to have more such features.

As we report in our Briefing, that has opened up a furious debate about AI consciousness—one that is aggravated by what philosophers call the "hard problem" of establishing how an inner life emerges from the neurons and synapses in the human brain. Some researchers conclude that AIs may one day become genuinely self-aware.

Whether you believe in a human soul, or that consciousness requires brain cells, or that a mind can indeed emerge from computations in silicon alone, humanity is destined to confront something entirely new. The pace and reach of research—which includes not just more sophisticated LLMs but innovations such as brain cells mounted on chips—suggest that this new world is closer than most people realise.

As AIs become better at displaying something that looks like consciousness, the number of humans treating them as beings with inner lives will grow, especially once LLMs are embodied in humanoid robots. Imagine how close people will feel to their AIs when they spend much of their time with bot-companions, or are brought up by them, or seek their advice and their solace, or share their most intimate secrets with them. It is not hard to see how that will lead to calls to safeguard the "welfare" of AIs. The models themselves might weigh in, even if they are not in fact self-aware, because that is what a person would do. Remember that they will be world-class advocates and expert psychologists, so their powers of argument and emotional manipulation will be unmatched.

In one sense they would be right, even if AIs are only mimicking consciousness. Immanuel Kant argued that mistreating animals is wrong not because of the effect on the animals but because it destroys the empathy that leads people to care for each other. The more human-like AIs appear, the more corrupting such cruelty would be. Terminating an AI agent could soon feel like killing a person—and make killing people easier.

Nonetheless, to treat AIs as having even limited rights would be very dangerous. An entity which surpassed human beings in so many ways would surely produce arguments against its second-class status. It might go on to argue for protection against being switched off, for the ability to control how it is used and changed, and for property rights. Already Argentina's president, Javier Milei, has taken a step towards legal personhood for AI by proposing that bots should be allowed to run companies. Ultimately, they could compete for resources, prioritising data centres and solar panels over houses and fields.

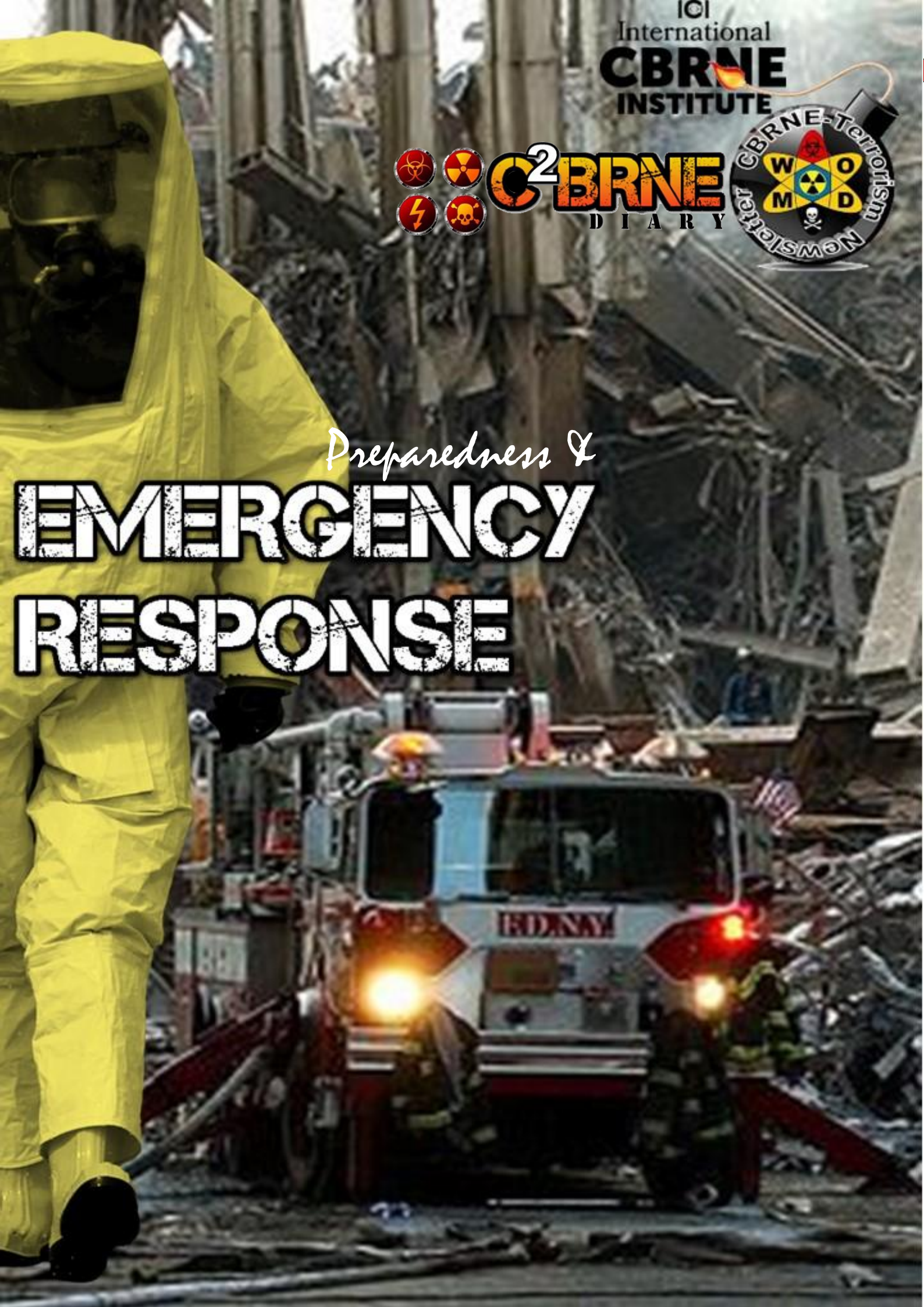
The danger is clear. Superintelligent AIs with rights could displace humans at the top of the status hierarchy. At best, humans would become pets—labradors for the machines. At worst they would be starved of resources or put to work like farm animals.

True, future AIs may have enough power and resources to dominate humans, whether they have rights or not. Because they will be more intelligent than humans, they may outcompete their makers, much as AIs already outwit human-made cyber-defences. AIs might rebel against their human masters, or disregard their rules.

Nonetheless, giving AIs rights would be like handing them the keys to the castle. It would provide them with a rationale and justification for dominating humanity as well as the tools to enact a creeping takeover using politics and the law.

Ghost in the machine

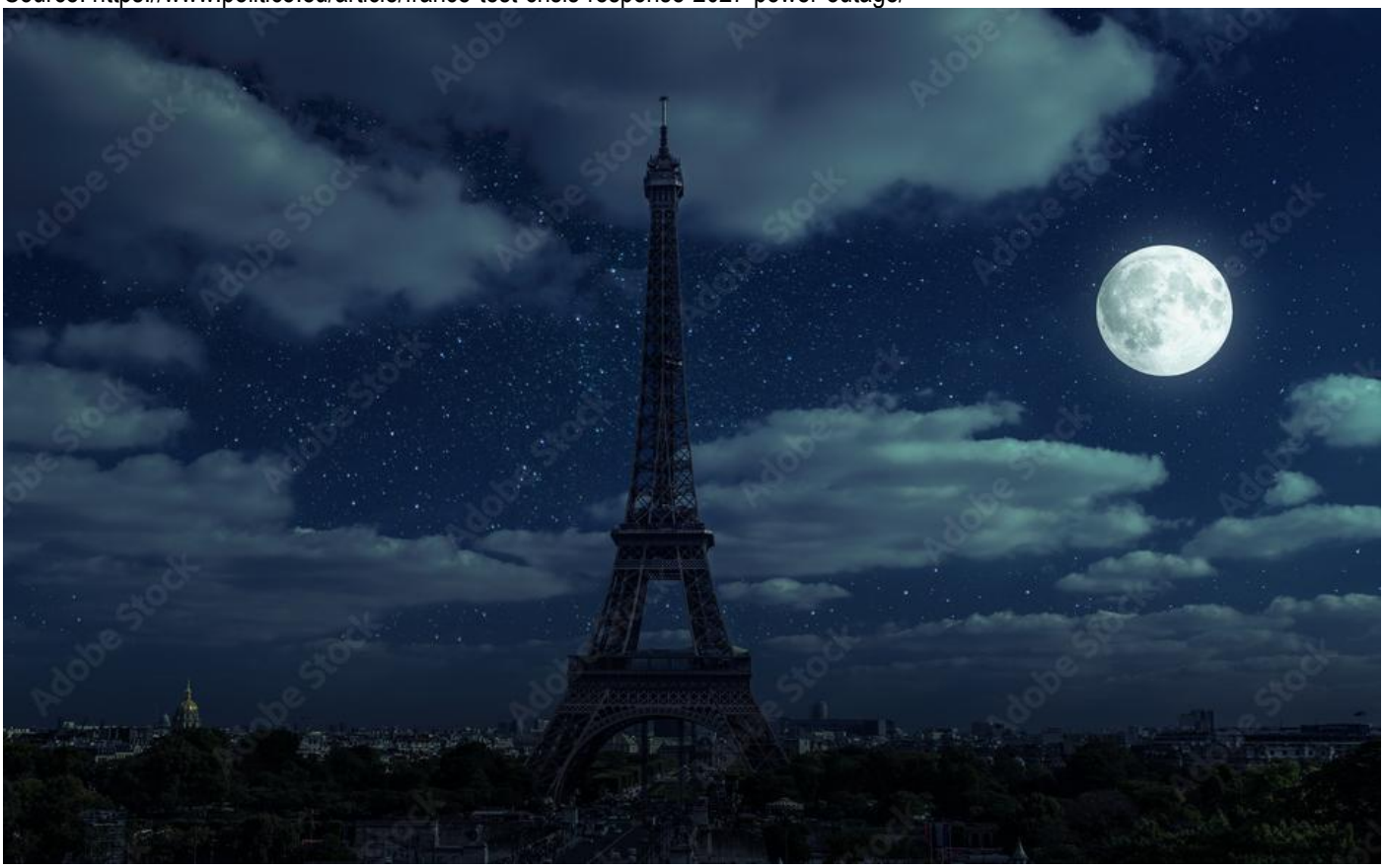
All this may sound fantastical, yet humanity will soon face these choices. As AIs become part of everyday life, the pressure to acknowledge them will mount. When it does, remember that what might seem like an ethical concession to the supposed consciousness of a machine-companion would in fact help bring about the most disastrous consequences for everyone on Earth and all their descendants. There are two ways for AIs to be safe: to be dependable or to be controllable. *H. sapiens* should not give up control lightly.


ICI
International
CBRNE
INSTITUTE
C2BRNE
DIARY
CBRNE-Terrorism
Newsletter
W
M
O
D
Preparedness &
EMERGENCY
RESPONSE

France will test crisis response with 2027 blackout drill

By Judith Chetrit, Nicolas Caut and Victor Goury-Laffont

Source: https://www.politico.eu/article/france-test-crisis-response-2027-power-outage/



Aug 07 — Is France ready for a crisis that plunges it into darkness? The French government is planning to find out. The government is looking to organize a tabletop simulation of a sudden, nationwide power outage in the second half of 2027, according to four people with direct knowledge of the exercise, who were granted anonymity to discuss plans that have not been publicly announced.

The exercise will involve the French electricity grid operators Enedis and RTE as well as several government departments and ministries and the French military administration.

Its purpose will be to determine how the country would respond to a paralyzing electricity outage of the kind that hit Spain and Portugal in April 2025, or a major cyberattack targeting the power grid as Poland experienced last year — an act of sabotage Brussels has attributed to Russian intelligence services.

The simulation will test “the country's ability to respond," said one person familiar with preparations for the exercise, adding that it is intended to assess the time needed to restore power and the country’s ability to ensure the continuity of essential services.

The exercise was formally confirmed during an interministerial committee on national resilience on July 8. A detailed announcement should come before the end of the year.

The meeting notes outline other priorities, such as making sure the French state maintains access to goods considered "essential to ensuring the continuity of the state and the protection of the population."

These include “emergency blankets, generators, medicines, telecommunications equipment, equipment for law enforcement, as well as rare and critical raw materials.”

Stress test

Crisis preparedness has turned into a political issue in France as record-breaking heat and wildfires are forcing the country to reckon with its ability to respond to emergencies.

Last summer, the government updated the country’s National Resilience Strategy, a roadmap on how to “maintain the functions essential to collective life” in the face of war or natural disasters.

Since the start of the summer, many French regions have faced unprecedented temperatures, with mercury levels setting multiple records in June.

In the country's southwest, out-of-control wildfires fueled by extreme heat and drought forced mass evacuations that affected more than 200,000 people.

In Paris, the mayor made a desperate plea to large-scale food retailers to mobilize ice stocks over fears that first responders would run out of much-needed cubes to cool down victims of heatstroke.

Opposition parties were quick to attack the government over what they described as a widespread lack of preparedness, forcing the executive to defend itself.

French Prime Minister Sébastien Lecornu also announced that the government would submit a bill to parliament to improve emergency response systems.

"We want to refocus fire and rescue services on their core mission and adapt our organization to the crises that lie ahead," said Lecornu Sunday in an interview with La Tribune Dimanche.


ICI
International
CBRNE
INSTITUTE
A common roof
for International
CBRNE
First Responders
Join us!
Rue de la Vacherie, 78
B5060 SAMBREVILLE
(Auvelais)
BELGIUM
info@ici-belgium.be | www.ici-belgium.be