

2

October 2026

10\26

CBRNE



EBOLA

Treatment Center



DIARY

*Dedicated to Global
First Responders*

ChatGPT-6

**Pharmacists
in crises**

**Visualize
bio-threats**

Plasma drones

AI feels pain

PART B

**Ebola
march**

**Plague
Siberia**

**Biohybrid
computers**

An International CBRNE Institute publication

IOI
International
CBRNE
INSTITUTE



DIRTY R-NEWS



Radiation Exposure May Trigger a Key Feature of Alzheimer's in The Brain

By David Nield

Source: <https://www.sciencealert.com/radiation-exposure-may-trigger-a-key-feature-of-alzheimers-in-the-brain>



Sep 27 – We're still searching for a root cause of [Alzheimer's disease](#), but scientists are learning more and more about the risk factors involved: old age, primarily, but also genetics, diet, sleep quality, and even pollution in the air we breathe.

Some of those risk factors, like our genetics and age, are set. Others are environmental factors – contributions to [Alzheimer's](#) risk from our surroundings – and exposure to ionizing radiation falls into this category.

Sources of radiation exposure and the evidence linking them to dementia risk. (Brenner et al., *Front. Immunol.*, 2026)

Medical (CT/Radiotherapy)

Most common exposure



Occupational (Nuclear/Cath Lab)

SELTIME: 80,348 workers;
SMR 1.29 uranium workers



Meta-analysis: RR 1.11 per
100 mSv (95% CI: 1.04-1.18;
P=0.001) n=18 studies



Environmental (Radon)

r=0.467 with AD mortality



Cosmic (Astronauts/Aviators)

Growing concern



Atomic Bomb Survivors

Methodological limitations

This is the radiation you get from working at nuclear sites, traveling through space, or (most likely for the majority of us) undergoing medical scanning procedures.

Ionization happens when radiation has enough energy to knock electrons off atoms, leaving charged ions that can damage the body.



In a new review published in [Frontiers in Immunology](#), researchers from The Feinstein Institutes for Medical Research in the US analyzed more than 100 previous papers to get a deeper understanding of how this radiation exposure might be linked to Alzheimer's.

"Human exposure to ionizing radiation is remarkably common," [write](#) the researchers in their published paper.

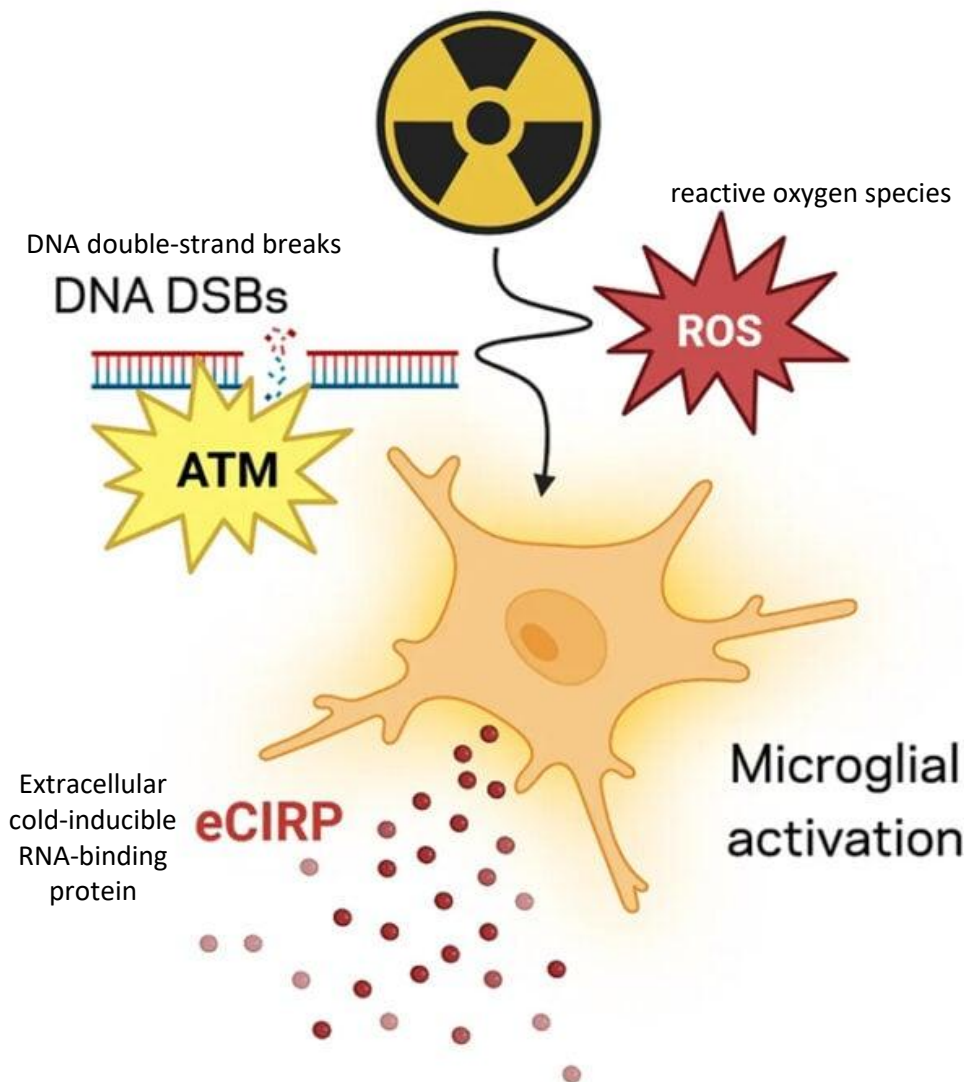
"The cognitive consequences of these exposures, particularly the potential for radiation to initiate or accelerate Alzheimer's disease, are therefore relevant to an expanding population."

The reviewed studies covered health and disease data, trials using animals, and molecular models, so there was a lot of information to bring together and parse.

For example, a [meta-analysis](#) of 18 studies found an 11 percent increase in dementia risk for every 100 millisieverts of cumulative radiation exposure. Some of the participants in these studies worked at nuclear plants and uranium processing facilities.

The hypothesis is that radiation triggers brain immune cells to release eCIRP, which then alters tau. (Brenner et al., *Front. Immunol.*, 2026)

Animal studies revealed similar patterns. Mice bred with Alzheimer's-like symptoms showed accelerated brain dysfunction when exposed to radiation. In monkeys, radiation was linked to damage to key



structural proteins in the brain.

"Available evidence suggests that ionizing radiation can induce a temporal sequence of molecular and cellular changes leading to cognitive decline," [write](#) the researchers.

The researchers also reviewed past work to develop ideas about how this radiation damage might occur. [An earlier study](#) from some of the same researchers established that damage done to DNA through ionizing radiation, and the subsequent brain immune cell response, leads to a release of a protein called eCIRP.

This then appears to drive a chemical change in tau proteins – phosphorylation – of the kind that precedes the toxic tangles seen in the brains of people with Alzheimer's.

While further research is going to be needed to confirm the connection and the chain reaction that goes on in human brains, it's a potential new target for Alzheimer's investigations.

"These structural and metabolic changes accumulate over time, shaping the clinical trajectory from radiation exposure to cognitive decline," [writes](#) the team.



"Translating these mechanistic insights into eCIRP-targeted therapies and validated biomarkers could open a new front in the prevention and treatment of these prevalent and debilitating disorders." According to the researchers, eCIRP could also be triggering inflammation and damage to the power supplies of cells, and putting restrictions on the creation of new neurons.

It's worth bearing in mind that the majority of the studies and data gathered here involve radiation doses far higher than most people would ever encounter day to day.

There's no suggestion that simply getting X-rays can lead to Alzheimer's, and it should also be noted that one of the researchers has registered a patent linked to eCIRP inhibition.

Taken as a whole, though, it's evidence that this is another avenue worth pursuing in Alzheimer's research. Next steps could include a closer look at the eCIRP-tau relationship and gathering new data at more realistic, lower radiation levels. "Dementia and Alzheimer's disease affect over 55 million people worldwide, and the number in the US alone is projected to reach 13.8 million, representing 3.8 percent of the population by 2060," [write](#) the researchers.

"Despite decades of research and the development of disease-modifying therapies targeting amyloid-beta and tau pathology, effective preventive and therapeutic approaches remain elusive."

► The research has been published in [Frontiers in Immunology](#).

Why the world's top nuclear nation is sitting out the hype on small reactors

By AnnaJulia Yates, Allison Macfarlane, and Alexander Wimmers

Source: <https://thebulletin.org/2026/09/why-the-worlds-top-nuclear-nation-is-sitting-out-the-hype-on-small-reactors/>



France's state-owned utility EDF announced plans to build 10 NUWARD small modular nuclear reactors (SMRs) in the European Union by 2035. But a closer look at France's broader energy strategy reveals a far more cautious reality. (Image: EDF)



Sep 25 – Many countries, including the United States, Canada, the United Kingdom, and Sweden, are gripped by a frenzy of enthusiasm for small modular reactors (SMRs). But France does not buy into the hype. Despite remaining staunchly pro-nuclear, with the world's largest share of electricity generation from nuclear reactors at nearly 70 percent, France gives SMRs only limited, conditional support, treating them as what they are: emerging technologies whose promised advantages have yet to be demonstrated at a commercial scale.

Varied enthusiasm for SMRs across countries may result from national industrial organization and history. The decentralized US power sector, for example, encourages [private development](#) of competing energy innovations, compounded by [over-budget and delayed large reactor projects](#) that have made SMRs seemingly more attractive. Conversely, France's state-led nuclear program, with its established supply chain and nuclear infrastructure, has focused on [large, standardized reactor builds](#). However, these differences between France and the United States should not be overstated; in Sweden, state-owned utility [Vattenfall is pushing for SMRs](#), and in the United Kingdom, the government is providing [substantial support](#) to development and potential deployment of SMRs.

But all countries face common fundamental problems with ageing fleets and issues bringing new reactors online. Average reactor ages are [39 years in France](#) and the United Kingdom, and 44 years in the United States and Sweden. France's most recent [Flamanville 3 reactor](#) took two decades to build, and at 6 times its original cost; Hinkley Point C in the United Kingdom has been repeatedly delayed at ballooning cost to consumers, and the most recent US newbuild projects resulted in the [vendor's bankruptcy](#). France's centralized system does not insulate it from the challenges SMRs claim to solve, making its cautious approach a noteworthy contrast to other countries' rush towards SMRs. It serves as a reminder that though

technologies provide easy political hype, they must prove safety, economic viability, and a clear advantage over existing low-carbon energy alternatives before major investment is justified.

A difficult start

Small modular reactors are promoted as a way out of nuclear power's most persistent problems of massive upfront capital requirements, cost overruns, and construction delays. Developers and [pro-SMR governments promise](#) that smaller, modularized designs will be cheaper to finance, easier to manufacture, and faster to build—among other [announced benefits](#).

The most obvious difference between SMRs and their larger counterparts is scale: The International Atomic Energy Agency (IAEA) defines an SMR as a reactor producing up to 300 megawatts of electricity—less than one third of what third-generation large reactors typically produce. The [OECD's Nuclear Energy Agency \(NEA\)](#) currently lists 127 SMR designs worldwide at various development stages.

[Despite hype and investment](#), the commercial record of SMRs remains thin, with only two reactors operational globally.

China [connected](#) its high-temperature gas-cooled SMR (called HTR-PM) to the grid in 2021. Russia's Akademik Lomonosov, a floating plant comprised of two pressurized water SMRs (KLT-40S), [has powered](#) the remote Arctic town of Pevek since 2019. In addition to these two reactors, only four demonstration projects are currently preparing or have started construction: GE-Hitachi's [BWRX-300](#) in Ontario, Canada, Kairos Power's [Hermes reactor](#) in Tennessee, TerraPower's [Natrium project](#) in Wyoming, and China National Nuclear Corporation's [Linglong One](#) in Hainan, China.

These projects do not prove the broader SMR case of efficient deployment: So far, most have faced delays, and the operating record is sparse. Available data from existing reactors show disappointing annual load



factors below [21 percent](#) for the Chinese HTR-PM in 2024 and around [30 percent](#) on average for the Russian floating reactor. (The load factor, or capacity factor, is a measure of reactor performance that reflects the actual electrical energy produced over a specific period compared to the maximum possible energy it would produce if it ran continuously at its reference design power. According to the IAEA, in 2023, the median load factor of all commercial power reactors worldwide was [88 percent](#).) A clearer sense of SMR cost, safety, and waste production will only emerge once more actual operational experience is gathered.

Another problem is the [heterogeneity](#) of proposed modular reactor designs, which creates uncertain safety criteria, yet-to-be-established supply chains, and novel waste streams. This mixed bag will further complicate—and slow down—SMR deployment by increasing regulatory, technical, and economic uncertainty.

Large reactors ‘avant tout.’

France relies on nuclear power for nearly [70 percent](#) of its electricity generation. With the addition of the [1650-megawatt-electric Flamanville 3 EPR](#) reactor, which was connected to the grid in 2024, the country now has [57 operable reactors](#). All of these are pressurized water reactors, with an average output capacity of around 1000 megawatts-electric.

France’s first commercial SMR project, named [NUWARD](#), began in 2017, when state-owned electric utility EDF partnered with the French Atomic Energy Commission (CEA) and two companies in which the French state owns major stakes: Naval Group and TechnicAtome. The [initial concept](#) promised a novel plant design consisting of two 170-megawatt-electric modules that would “serve the global market with a competitive solution in the late 2020s.” But in 2024, NUWARD [abandoned that design](#) due to technical and market-viability issues, deciding instead to “pivot its SMR product strategy” toward “a design based on proven technology bricks

only.” EDF [later announced](#) that the new design would be a much larger reactor whose 400 megawatt-electric capacity exceeds the traditional 300 megawatt-electric cutoff for SMRs.

The only other significant French SMR initiative was launched under President Emmanuel Macron’s 2030 investment plan. Published in 2021, the plan includes [€129.8 million in state support](#) for 11 advanced nuclear projects, some of which are SMRs. But this support remains conditional and relatively restrained: Participants will have to complete the standard licensing protocol to ensure technical and safety credibility and continually convince the state of their projects’ economic viability before accessing future rounds of funding.

One reason for France’s cautious approach to SMRs is that it already has a well-defined, [centralized](#), and [standardized](#) nuclear program, supported by decades of investment. The French government is directly involved in setting national energy priorities, tasking EDF with constructing and “[frenchifying](#)” US-designed large pressurized water reactors and refining each successive reactor generation, ultimately standardizing its nuclear fleet.

The clearest expression of this strategy is France’s prioritization of the large EPR2 reactor—a modernized version of the EPR reactor now operating in France, Finland, and China—over SMRs. In 2020, *Nuclear Intelligence Weekly* [reported](#) that EDF “remains almost exclusively focused on developing a more economic version of the EPR”, and focused worker capacity towards the EPR2 with “as many as 1000 engineers involved” in the project versus only “10 to 15” on NUWARD development.

President Macron made this priority explicit: In 2022, he announced plans to build [six new EPR2 units](#) once the design is ready, giving large reactors a defined place in the future French nuclear fleet that SMRs have yet to receive. Although the EPR2 design remains [conceptual](#) and the first operating reactors are [not expected before 2040](#),



Macron's announcement and EDF's staffing choices suggest that SMRs remain secondary to France's long-held strategy of building large reactors.

Skepticism of SMRs

Beyond caution, diverse actors inside the French nuclear establishment have raised concerns about SMRs. French nuclear safety authorities, for instance, have been vocal about the feasibility and the safety issues of certain SMR approaches. In 2021, the [French nuclear regulator ASN warned](#) that many SMR designs rely on innovative solutions whose "feasibility and efficiency are yet to be demonstrated." A 2023 parliamentary assessment [raised similar concerns](#), noting that SMR projects show "varying maturities" and that "important research and development efforts remain to be done" for some technologies, including molten-salt, lead-cooled, sodium-cooled, and high-temperature SMRs. In May, the French regulator [added](#) that proposed SMR deployment near industrial or densely populated areas would require adapted safety objectives. France's major energy union FNME-CGT has been equally critical of the private governance of SMR development. The organization argues that the public interest, which the nuclear energy sector ought to cater to, is at odds with the short-term interests of private start-up companies. [In a scathing response](#) to one of the selected projects in the 2030 reactor investment program, the union wrote: "the [SMR] project ... is the work of a 'start-up' whose shareholding is, to say the least, capitalist, and raises technical and environmental questions." The union reiterated that advanced reactor projects ought not to be handed to private companies, and "even less so" to start-ups, while voicing concerns about the environmental, territorial, labor, security, and budget risks of advanced reactor programs. Yves Bréchet, a former High-Commissioner for Atomic Energy of France, has offered a broader warning about the political logic driving SMR enthusiasm. [Testifying before the Belgian parliament in](#)

[2023](#), he cautioned: "SMRs must not become a pretext for dispersing resources in pursuit of political optics." His concern was that government funding for SMRs may be more in pursuit of press releases about technological ambition and international competitiveness than concrete potential. Any decision to pursue them, Bréchet continued, must come from a "thorough scientific and technical review" by an expert body, not from "arbitrary [political] decision."

Lessons for SMR advocates

France's caution about SMRs should raise flags for policymakers and nuclear organizations contemplating these reactors elsewhere. The country's prioritization of large reactors and the skepticism voiced by some French experts point to a clear lesson for countries considering them.

Reactor programs should not be based on hype or political spectacle: A privatized, start-up-led, and lightly regulated approach may generate momentum and attract investors, but it cannot be a substitute for technically credible proof that a reactor is safe, economical, and able to withstand institutional scrutiny. Any policy decision to champion SMR development must be justified by evidence that they are not only technically, commercially, and politically viable, but also necessary: able to address a demonstrated energy need better than traditional large reactors or other available alternatives, like renewables, on time and aligned with societal and environmental objectives. Especially given limited budgets and workforce capacity, governments cannot afford to chase every potentially promising design at once. SMRs cannot be immune to these constraints.

Betting on SMRs will mean choosing to allocate engineers, regulators, public money, and political attention away from other energy options that can deliver low-carbon energy now instead of sometime in an uncertain future. Enthusiasm and innovation alone cannot outweigh the need for concrete evidence for SMRs.



AnnaJulia Yates is a research assistant at the School of Public Policy and Global Affairs at the University of British Columbia.

Allison M. Macfarlane is professor and director of the School of Public Policy and Global Affairs at the University of British Columbia. She chaired the US Nuclear Regulatory Commission from 2012 to 2014. Macfarlane holds a doctorate in geology from MIT and a bachelor in science from the University of Rochester. Macfarlane served on the White House Blue Ribbon Commission on America's Nuclear Future.

Alexander Wimmers is a postdoctoral research associate at the Workgroup for Infrastructure Policy at Technische Universität (TU) Berlin, a guest researcher at the German Institute for Economic Research (DIW Berlin), and a visiting faculty member of the School of Public Policy and Global Affairs at the University of British Columbia. He holds an M.Sc. in industrial engineering from RWTH Aachen and a doctorate in economics from TU Berlin.

الجرعة المنخفضة... أين يبدأ الخطر؟

من 1 mGy إلى 1 Gy ماذا نعرف؟ وماذا لا نعرف؟

ما الذي نعرفه بالجرعة المنخفضة؟

الجرعة المنخفضة ليست رقماً واحداً ثابتاً، بل هي نطاق يختلف باختلاف:

- نوع الإشعاع (ألفا، بيتا، غاما، نيوترونات...)
- الهدف من التعرض (تشخيص، علاج، مهني...)
- العضو أو النسيج المعرض.
- معدل الجرعة ومدة التعرض.

الجرعة المنخفضة لا تعني صفر خطر، كما أن وجود خطر محتمل لا يعني أن التأثير حتمي.

الجرعة: كمية الطاقة الممتصة

الجرعة الممتصة (Absorbed Dose)

mGy (ملي جراي)

كمية الطاقة الممتصة في وحدة الكتلة من السليخ.

الجرعة الفعالة (Effective Dose)

mSv (ملي سيفرت)

تأخذ في الاعتبار حساسية الأنسجة المختلفة، وتستخدم لتقدير الخطر الإشعاعي الكلي.

مفهوم مهم

الجرعة المنخفضة (وفقاً لـ UNSCEAR) هي أقل من:

100 mGy

كما أن معدل الجرعة المنخفضة يُعرف بأنه $\leq 0.1 \text{ mGy/min}$ (أي $\leq 6 \text{ mGy/h}$).

هذه قيم وصفية علمية وليست حدود أمان تنظيمية.

نطاقات الجرعة ومستويات الخطر

< 10 mGy	10 – 100 mGy	100 mGy – 1 Gy	> 1 Gy	1 Sv وما فوق
منخفض جداً	منخفض	متوسط	مرتفع	جرعات شديدة
لا يُتوقع حدوث تأثيرات نسيجية حادة. تقدير الخطر السرطاني صعب وباتياً. الخطر المحتمل: ضئيل جداً.	لا توجد عادة تأثيرات نسيجية حادة. الخطر السرطاني - إن وُجد - صغير ويصعب قياسه بدقة. الخطر المحتمل: صغير.	تبدأ الأدلة الوبائية على زيادة الخطر مع ارتفاع الجرعة. الخطر المحتمل: يزداد تدريجياً.	تبدأ التأثيرات النسيجية (مثل احمرار الجلد، تساقط الشعر، إنبات عضة العين...) في تصبح ذات أهمية متزايدة. الخطر المحتمل: مرتفع.	يمكن أن تظهر متلازمة الإشعاع الحادة. وتزداد شدة التأثير على العضو ومعدل الجرعة. الخطر المحتمل: كبير جداً.

ماذا نعرف عن السرطان عند الجرعات المنخفضة؟

استخدم هيثم الوقاية لنموذج الخطي بدون عتبة (LNT). لتقدير الخطر عند الجرعات المنخفضة، وهو نموذج وقائي وليس دليلاً مباشراً. من الدراسات الوبائية:

تزداد صعوبة إثبات زيادة الخطر بسرطان مع انخفاض الجرعة. والنتيجة هي تقديرات مستندة إلى الأدلة المتاحة.

أمثلة على الجرعات في الحياة اليومية

- أشعة سنية (Dental X-ray) $\approx 0.005 - 0.02 \text{ mGy}$
- أشعة الصدر (Chest X-ray) $\approx 0.02 - 0.1 \text{ mGy}$
- الأشعة المقطعية للجسم كاملاً (CT) $\approx 10 - 100 \text{ mGy}$ (تتباين حسب الإجراء والمنطقة)
- المعدل السنوي من الخلفية الطبيعية (في بعض المناطق) $\approx 2.3 \text{ mSv}$

ما الذي يحدد الخطر؟

- الجرعة (mGy / mSv)
- معدل الجرعة
- نوع الإشعاع
- العضو أو النسيج المعرض
- العمر والحالة الصحية

الجرعة ليست الرقم .. بل السياق الذي يعطيها معناها

د. عمر دسوقي
أستاذ الفيزياء الحيوية الإشعاعية

MIRA Safety Launches Free Nuclear Blast Simulator to Help the Public Understand and Prepare for Nuclear Emergencies

Source: <https://www.mirasafety.com/pages/nuclear-simulator>

Oct 05 – MIRA Safety® has announced the launch of its [Nuclear Blast Simulator](#), a free, interactive nuke map available on the MIRA Safety website. The tool allows users to visualize the potential blast radius, fallout map and recommended safe shelter time for a nuclear or radiological event, giving individuals, families and organizations a clear picture of how such an event could affect their area and what steps they can take to protect themselves.

NUCLEAR SCENARIO · EDUCATIONAL TOOL

Nuclear Blast & Fallout Simulator

SCENARIO SETUP

Configure effects

Ground zero

City, ZIP, address, or lat, lon

Search, enter latitude and longitude, or click the map.

Weapon preset

W87-0 · Minuteman III · 300 kt

United States Current Public · High Confidence

Mk21/SERV warhead currently carried by Minuteman III.

Source: Federation of American Scientists / Bulletin of the Atomic Scientists

Public, non-operational data only. Estimated entries require independent SME and editorial approval.

Burst type

Airburst **Surface burst**

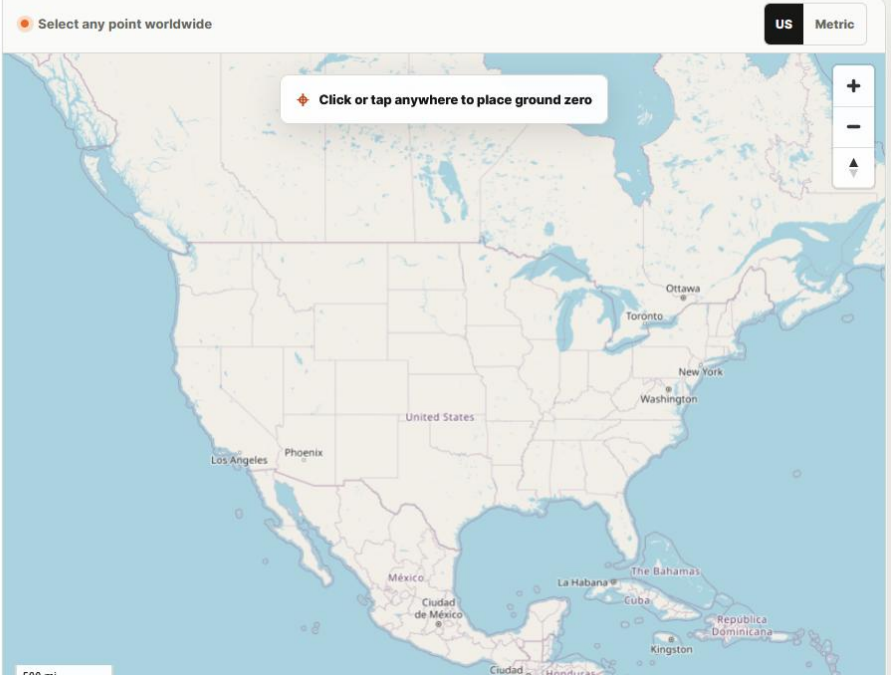
Advanced options

Select ground zero to continue

Select any point worldwide

Click or tap anywhere to place ground zero

US Metric



© OpenStreetMap contributors

Educational estimate. Simplified public models; not a real-time emergency forecast.

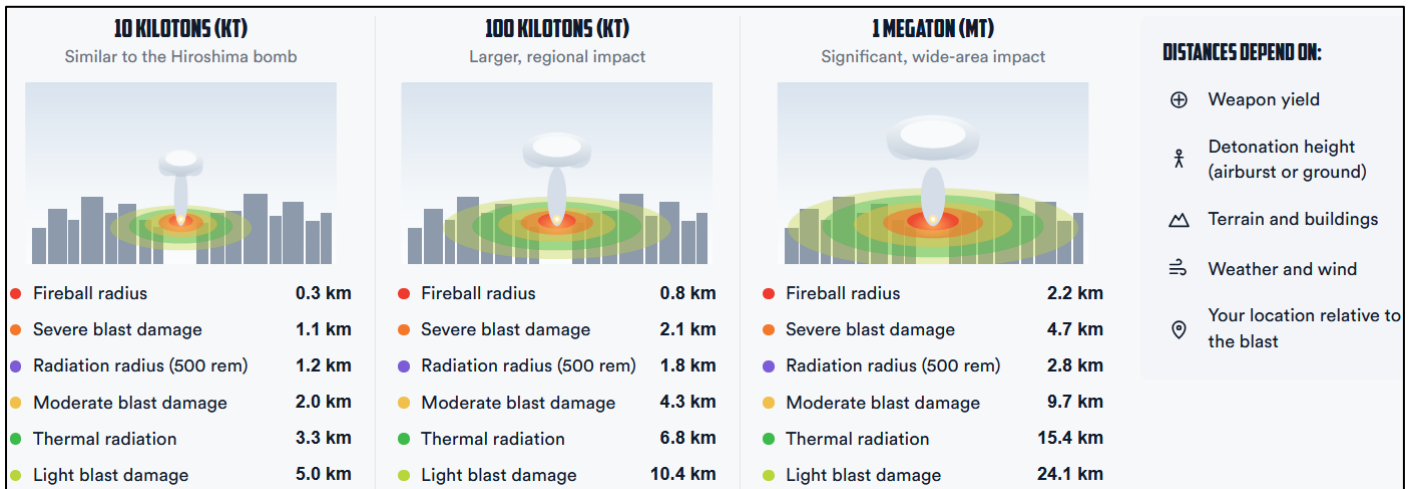
"Most people have never seriously thought about what a nuclear event would mean for their own neighborhood, and that uncertainty is what leaves them unprepared," said [Roman Zrazhevskiy](#), Founder and Chief Executive Officer at MIRA Safety. "The Nuclear Blast Simulator replaces fear and guesswork with information. When you can see where the blast zones fall, where fallout could travel and how long you would need to stay sheltered, preparedness stops being abstract and becomes a plan you can act on."

Nuclear and radiological emergencies are among the least understood threats, and in the first hours after a detonation, decisions about where to shelter and for how long have the greatest impact on survival. The Nuclear Blast Simulator helps close this knowledge gap by showing users how distance, fallout patterns and shelter timing shape their risk, so they can make informed decisions before an emergency rather than during one.

Preparedness, however, goes beyond knowing the risk. MIRA Safety offers a full line of professional-grade CBRN protective equipment to support every stage of a nuclear emergency plan, including full-face respirators such as the [CM-8M mask](#), CBRN filters rated against radioactive particles, protective suits, and radiation detection devices. Together with the simulator, these



solutions give civilians, first responders and organizations both the knowledge and the gear they need to shelter safely and reduce exposure to radioactive fallout.



"Our mission has always been to make advanced protection accessible to everyone," added Zrazhevskiy. "Education is the first layer of that protection. The simulator helps people understand the threat, and our equipment helps them respond to it with confidence."

The Nuclear Blast Simulator is available now and free to use. **To try the tool**, visit the [MIRA Safety Nuke Map](#), or explore protective equipment on [MIRA Safety's e-commerce platform](#).

About MIRA Safety: MIRA Safety is a leading global provider of professional-grade personal protective equipment, trusted by military, law enforcement, industrial, and civilian users worldwide.

Georgia arrests man over attempted \$700,000 uranium sale

Source: <https://www.reuters.com/3fec1b0a85a0/world/us/georgia-arrests-man-over-attempted-700000-uranium-sale-2026-10-07/>

Oct 07 – Georgia has detained a man who planned to sell uranium-238 to a foreign national for \$700,000, the national security service said on Wednesday.

The Georgian citizen is accused of illegally handling nuclear material and faces up to 10 years in prison, it said in a statement, sharing a video of armed security personnel pulling a man out of a white sedan and handcuffing him on the ground.

It was not immediately possible to contact the man or a legal representative for comment.

The footage also showed agents testing two small containers in the trunk of the vehicle with a handheld radiation detector. One of them appeared to contain a light-coloured powder.

Uranium-238 is the most common uranium isotope and, while radioactive, is not ordinarily usable on its own to make a nuclear bomb.

The security of nuclear materials was one of the biggest concerns after the 1991 fall of the Soviet Union, of which Georgia was part. There have been several serious incidents involving the illicit trade in nuclear materials in Georgia over recent decades.

In February, Georgia detained two people [who attempted to purchase \\$3 million worth of uranium](#) and a cache of a radioactive isotope found in nuclear weapons testing programmes.

EDITOR'S COMMENT: Georgia has witnessed several notable cases of radioactive material smuggling. In October 2025, three Chinese nationals were



arrested in Tbilisi for allegedly plotting to purchase 2 kilograms of uranium for \$400,000 and smuggle it to China via Russia. In February 2026, two foreign nationals were detained in Kutaisi for attempting to buy uranium and cesium-137 worth \$3 million, highlighting Georgia's ongoing role as a transit point for nuclear contraband following the Soviet Union's collapse. Several incidents have been recorded before 2025: In June 2003 (Armenian Border): Authorities seized 170 grams of highly enriched uranium (HEU) enriched to 90%; in February 2006 (Tbilisi): A Russian national was arrested in a sting operation at the Gldani Market with about 100 grams of HEU (89% enrichment) concealed in a plastic bag; in March 2010: Two Armenian citizens were arrested for transporting 17 grams of HEU in a pack of cigarettes; in 2012 (Multiple Locations): Several cases involved Cesium-137, including 700 grams seized in Tbilisi from research reactor filters, and smaller amounts in Batumi and Samtredia; in 2016: Authorities arrested 121 people in two sting operations for attempting to sell uranium-238 and uranium-235; in July 2019: A batch of Thorium-232 (71.63 kg) was intercepted at the Sadakhlo border checkpoint near Armenia; in 2023 (Poti): A man was arrested for attempting to sell a small bottle of uranium for \$2 million.

PERSPECTIVE
Open Access


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
Rump et al. Military Medical Research (2018) 5:27

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



Russia Modernizes Secret Facility 1335 Linked to Dead Hand Automated Missile Launch System

Source: <https://militarnyi.com/en/news/russia-modernizes-secret-facility-1335-linked-to-dead-hand-automated-missile-launch-system/>



One of the entrances to the underground complex within Mount Kosvinsky Kamen—where Object 1335 is located—is connected to the "Perimeter" system. Photo: Danwatch

Sept 27 – Russia is upgrading a secret underground facility inside Mount Kosvinsky Kamen in the Urals, reportedly linked to the Perimeter nuclear command-and-control system.

The findings come from an investigation by [Danwatch](#), conducted in cooperation with [The Sunday Times](#) and the Dossier Center.

Known as Facility 1335, or by its codename “Grot,” the facility is located inside Mount Koscinsky Kamen in the Sverdlovsk region.

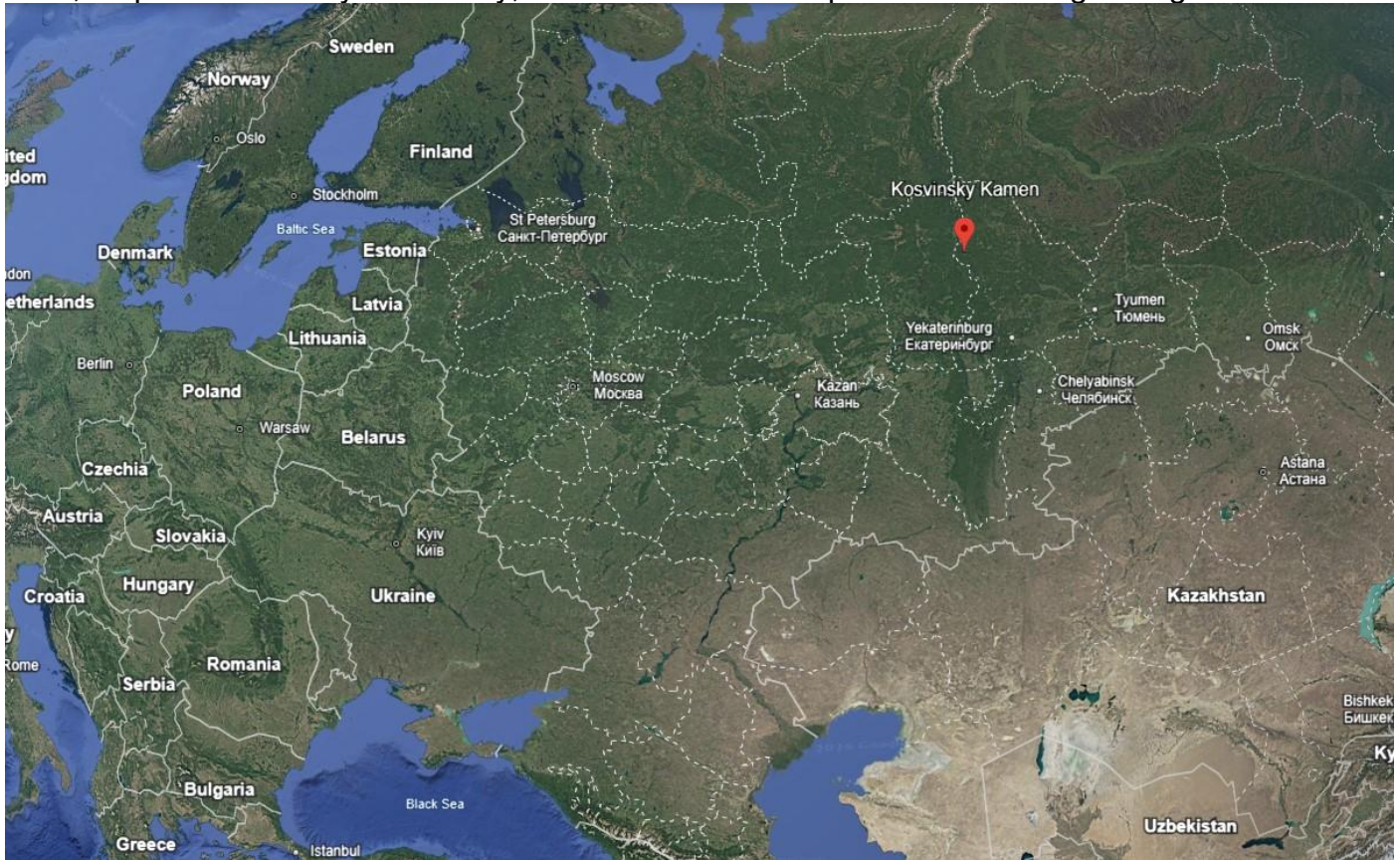
Russia’s Ministry of Defense began major modernization work on the facility in 2018.

The investigation found thousands of government defense procurement contracts related to Facility 1335 between 2018 and 2024.

These include at least 22,000 tonnes of steel, more than 56,500 tonnes of cement, around 600 kilometers of electrical cable, 21 electrical substations, 182 ventilation



units, 95 pieces of heavy machinery, and hundreds of components for drilling through rock.



Mount Kosvinsky Kamen. Satellite image from Google Earth



Separately, the Russian military procured specialized ventilation systems, filters designed to protect against radioactive dust, blast-resistant equipment, and communications systems.



The scale of the work goes far beyond routine repairs. The documents refer to at least 24 underground rooms, for which a compressed-air system was installed. The distance between the facility's main and emergency exits is about 2.5 kilometers.

In 2019, refrigeration equipment with a total capacity of around 5 MW was purchased for Facility 1335. It removes heat generated by communications, computing, and electrical equipment that may need to operate continuously.

A separate part of the modernization effort focuses on protecting and sustaining the facility. The documents mention blast-resistant structures, engineering equipment, and other systems required to operate a hardened command post.

The researchers also identified purchases of Western-made equipment, including Rockwool insulation, Danfoss fittings, Belimo valves, and other products.

Satellite imagery also captured extensive earthworks near the facility's entrances. According to [Danwatch](#), the Russian military was constructing new entrances into the mountain and carrying out work inside the underground system.

The Perimeter System, or Dead Hand

The Perimeter system, often known in the West as Dead Hand, is a Soviet-era system that Russia still uses as a backup command system for its strategic nuclear forces.

It was developed during the Cold War to ensure that Russia could still launch a retaliatory nuclear strike even if its main command centers were destroyed and communication with the country's top military and political leaders was cut off.

The system consists of hardened command centers, secure communication networks, and special command missiles. Once launched, these missiles transmit coded orders to strategic nuclear missile units, including those that may have lost their normal communication links.

The system is designed to preserve Russia's ability to launch a retaliatory nuclear strike even after a nuclear attack has disrupted the country's central command and communications.

Meanwhile, details about how the Perimeter system actually operates remain largely classified. Therefore, the widely circulated claim that the system can independently and automatically launch Russia's entire nuclear arsenal without human involvement needs to be qualified.

According to the [Dossier Center](#), a new component of the system, developed as part of its modernization, was installed at the facility as early as the 2000s. Experts believe the current work aims to maintain and upgrade the facility's operational capabilities.

Retired Major General Viktor Khalin, who helped develop the system during the Soviet era and oversaw its modernization from 2005 to 2010, has also written about Perimeter's modernization.

According to Khalin, after the upgrade, the system could operate not only in an automated mode but also in a fully automatic mode.



How do we transmit a radioactive source without allowing its radiation to reach those around it?

By Dr. Omar Desouki | Professor of Radioactive Biophysics | Egypt

Source: <https://www.linkedin.com/feed/update/urn:li:activity:7517181772447719425/>

سلسلة: نقل المواد المشعة

كيف تُصمَّم عبوات نقل المواد المشعة؟

هندسة دقيقة ... لحماية الإنسان والبيئة

1 مكونات العبوة ودورها

تصمَّم عبوات نقل المواد المشعة وفق متطلبات هندسية صرامة تضمن احتواء المادة المشعة وتوفير التدريع الكافي، ومقاومة الظروف الطبيعية والحوادث المحتملة أثناء النقل.



الرصاص
(ممتاز لطاقت غاما لمتوسطة والعالية)



التنجستن
(كثافة عالية، وجسم أقل)



نقطة الرفع والتثبيت
للتعامل الآمن مع العبوة أثناء النقل.

الغلاف الخارجي
عادة من الفولاذ المتين مقاوم للصدمات والحرارة.

التدريع الإشعاعي
مثل الرصاص أو التنجستن لتقليل شدة الإشعاع النافذ.

العبوة الداخلية
تحتوي على المادة المشعة في حاوية محكمة الإغلاق (مصنوعة من الفولاذ المقاوم للصدأ).

قاعدة العبوة
لتثبيت العبوة في وسيلة النقل.

2 القانون الفيزيائي المستخدم

قانون التوهين الأسّي لأشعة غاما

$$I = I_0 e^{-\mu x}$$

I : شدة الإشعاع بعد المرور في المادة.
 I_0 : شدة الإشعاع قبل المرور (عند المصدر).
 μ : معامل التوهين الخطي للمادة.
 (يتحدد بنوع المادة و طاقة الإشعاع).
 x : سمك مادة التدريع.



المصدر الإشعاعي (جاما) مادة التدريع (رصاص / تنجستن) الكاشف الإشعاعي

ملاحظة مهمة

• هذا القانون يمثل نموذجاً مبسطاً للحزمة الضيقة من الإشعاع. في العبوات الحقيقية، يؤثر الإشعاع المتشتت وتراكم الفوتونات على الجرعة خلف التدريع. ولذلك يتطلب التصميم الفعلي حسابات أكثر اكتمالاً واختبارات منهجية.

عبوة مصممة جيداً ... نقل آمن

د. عمر دسوقي
أستاذ الفيزياء الحيوية الإشعاعية

Tiny Specks of Chernobyl's Nuclear Fuel Are Still Intact After 40 Years

Source: <https://www.sciencealert.com/after-40-years-chnobyls-radioactive-particles-are-surprisingly-stable>

Oct 17 – We're now 40 years on from [the worst nuclear disaster](#) in history, and scientists have been analyzing radioactive particles blown out by Chernobyl to see what's left of the fuel that got scattered – and what state those particles are in.

Knowing how these chemical pellets have held up to decades of weathering and decay is

radioactive dust particles, which showed much more stability than expected.

That means most of the radioactive material remains locked inside.

The researchers behind the work, from Leibniz University Hannover and the Helmholtz-Zentrum Dresden-Rossendorf (HZDR) in Germany, have published their findings in the [Journal of Hazardous Materials](#).

Six fuel particles were analyzed in total, split into three categories depending on what happened to them during the original reactor explosion. Some, for example, had melted and bonded with their coating of zirconium.

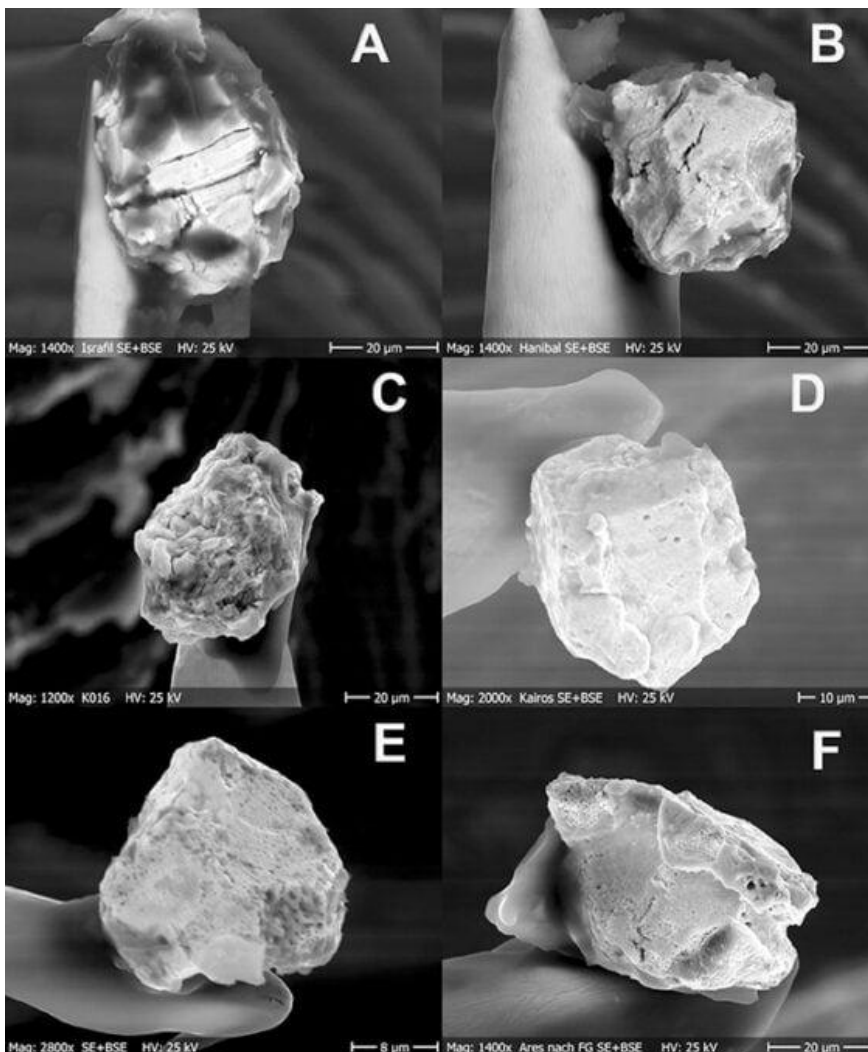
"There are three classes of these particles," [says](#) physicist Tobias Weissenborn, from Leibniz University Hannover.

"First, there are particles that are chemically and physically still very similar to the nuclear fuel uranium dioxide.

"Then, there are particles that are partially or fully encased in, or completely fused with, their zirconium layer. The third type of particle was created when the graphite moderator caught fire."

The researchers studied six different radioactive particles recovered from Chernobyl. (Weissenborn et al., *J. Hazard. Mater.*, 2026)

Each of these particle classes has ended up with a different chemical composition, which can affect how they degrade. The sampled particles measured a mere 8–50 micrometers in size (so at least 20 would fit inside a single millimeter), and were all collected from inside the [Chernobyl Exclusion Zone](#).



valuable for understanding the safety of nuclear plants in the future – and for working out when it might be safe to return to the Chernobyl site without a protective suit.

What stood out most from the analysis was the ["remarkably persistent"](#) structure of the



A technique known as diffraction analysis was used to look inside these microscopic specks, with the atomic crystal structure being determined through the way X-rays reflected away from it.

"Ultimately, we fully rotated the particles in an X-ray beam focused down to 100 micrometers – about the thickness of a human hair," [says](#) crystallographer Christoph Hennig from HZDR.

"We measured each particle from 2,000 different angles to seamlessly capture all the reflections."

Across all three particle types, the researchers found the structure holding the nuclear fuel had largely remained intact. That suggests a greater chemical stability in these fuel specks than previously assumed.

The study also gives scientists a better idea of how these particle groups are put together – exactly how the disaster and its subsequent chemical chain reaction shifted the composition of the fuel crystals.

"Our findings demonstrate that Chernobyl fuel particles act as remarkably persistent reservoirs," [write](#) the researchers in their published paper.

frameborder="0" allow="accelerometer; autoplay; clipboard-write; encrypted-media; gyroscope; picture-in-picture; web-share"

referrerpolicy="strict-origin-when-cross-origin" allowfullscreen>

The researchers frame this as potentially good news: It means the particles may be releasing radioactive material into the environment more slowly than previously assumed.

While it's not safe to go back, the findings could help scientists better assess the long-term environmental risks.

However, there's a flip side.

The radioactive material hasn't gone anywhere, which means these particles will remain dangerous for decades. Breathing these particles in is still going to be extremely hazardous to health, and they could get kicked up from the ground by wind, erosion, or passing animals.

The researchers also acknowledge that six particles – all sourced from two locations – are a small sample. Weighing up the current risks in Chernobyl will require more comprehensive sets of data collected over longer periods.

"Even if we obtained some averages at some point, we still wouldn't be able to make universal statements about health risks in the region," [says](#) Weissenborn.

"Even if the particles decay in a largely uniform pattern, there will always be outliers, more persistent particles, that will release radionuclides at a later point in time."

●► The research has been published in the [Journal of Hazardous Materials](#).

Beyond nuclear fatalism: The UAE paradigm

By Aysha Taryam | Editor-in-Chief, Gulf Today News and Media.

Source: <https://www.gulftoday.ae/opinion/2026/10/18/beyond-nuclear-fatalism-the-uae-answers-einsteins-call>

Oct 18 – For decades, the global conversation surrounding nuclear security has been held hostage by a cynical premise: that true safety can only be bought by the threat of total annihilation. Strategic orthodoxy has long taught that the Cold War doctrine of Mutually Assured Destruction (MAD) is the ultimate guardian of global stability, creating a terrifying calculus where peace is preserved not by goodwill, but by the shared terror of the abyss.

Yet, as geopolitical tensions fray and nuclear risks climb to levels unseen in a generation, this logic looks less like a shield and more like a slow-fuse trap. True statecraft requires a higher ambition. It demands that we look beyond the balance of terror and return to an older, more courageous premise: that security built on the capacity for indiscriminate devastation is inherently fragile.



Nearly a century ago, warning lights were already flashing. Even the brilliant minds who unlocked the secret of the atom were among the first to recognize its terrifying existential threat. When Albert Einstein and philosopher Bertrand Russell released their historic 1955 manifesto at the dawn of the thermonuclear age, they issued a stark, enduring challenge to humanity. Confronted with the prospect of universal death, Einstein famously urged that "we shall have to learn to think in a new way," reminding world leaders that their primary obligation was no longer to secure victory for a single nation, but to ensure the physical survival of the human race.

That philosophical call to action is too often sidestepped in the corridors of power, but it lives on in the active foreign policy architecture of the United Arab Emirates. Since its inception, the nation has carried the banner for the peaceful resolution of conflict. As the UAE's founding father, the late Sheikh Zayed Bin Sultan Al Nahyan, famously observed, the country's foreign policy is rooted in justice, peaceful dispute resolution, and the "absolute rejection of weapons that cause indiscriminate devastation to humanity." In doing so, the UAE operationalized Einstein's mandate, replacing nuclear fatalism with strategic sanity.

That foundational philosophy is not merely a dormant legacy; it is an active, evolving blueprint. At this month's United Nations General Assembly First Committee debate in New York, the [UAE once again placed this principle front and centre](#), reiterating its steadfast conviction that the only absolute guarantee against the use or threat of weapons of mass destruction is their total elimination.

To examine the philosophy of disarmament is to dismantle two major falsehoods that have long paralyzed international diplomacy.

The first falsehood is the cynical claim propagated by nations seeking to reign supreme, who weaponize the narrative of security to grant themselves permission to possess weapons of mass destruction in their quest for dominance. This logic assumes that global standing and national survival can only

be guaranteed through the capacity for total destruction.

By proving that a modern, prosperous state can rise without the shadow of the bomb, the UAE provides the exact moral and practical compass the world needs to step back from the brink and finally choose peace.

The United Arab Emirates shatters this paradigm. Rather than pursuing influence through threats of domination, the UAE has built its strategic power on active diplomacy, constructive peacekeeping, and regional partnership. To prove that a nation can achieve advanced modernization without crossing the nuclear threshold, the UAE turned vision into reality through the Barakah Nuclear Energy Plant. This landmark civil nuclear program serves as the ultimate proof that advanced scientific development and clean energy can coexist seamlessly with rigorous nuclear integrity. Rather than sheltering behind the threat of total destruction to enforce peace, the UAE chose uncompromising accountability. It ratified the IAEA Comprehensive Safeguards Agreement and Additional Protocol, and famously signed the "123 Agreement" with the United States, voluntarily forgoing domestic uranium enrichment and spent fuel reprocessing.

In doing so, the UAE established a vital philosophical truth: national sovereignty and long-term security do not rely on weapons of mass destruction. Maintaining robust conventional capabilities to defend against hostile threats is fundamentally different from acquiring weapons of mass destruction. True statecraft proves that a sovereign nation can firmly protect its borders and people against aggression while remaining resolutely opposed to the catastrophic shadow of WMDs. The second falsehood is the fatalistic belief that regional stability is best managed through an arms race or fragile nuclear standoffs. In international forums, the UAE consistently champions the establishment of a Middle East completely free of weapons of mass destruction, a vision that requires collective multilateral engagement, adherence to the



Treaty on the Non-Proliferation of Nuclear Weapons (NPT), and deep-seated diplomatic trust.

To truly grasp why this model matters today, we must look closer at the changing nature of modern conflict and state security. Traditional theories of deterrence were born of a bipolar world order that no longer exists. Today, proliferation risks intersect with cyber vulnerabilities, artificial intelligence, and regional proxy instabilities, exponentially raising the danger of catastrophic miscalculation. In this hyper-connected and fragile landscape, maintaining vast nuclear stockpiles is no longer a stabilizing anchor; it is a ticking clock.

The philosophy championed by the UAE reframes state security away from fear and toward resilience. True resilience is not measured by the destructive capacity a nation can wield against its neighbours, but by its ability to foster sustainable development,

diplomatic trust, and cooperative alliances. When nations invest their capital and intellect into clean energy, scientific innovation, and regional integration rather than atomic stockpiles, they elevate both their own stature and the collective security of humankind.

Ultimately, the choice facing global leaders is stark. We can continue down the well-worn path of nuclear fatalism, accepting the permanent Sword of Damocles as the price of doing business in international politics, or we can embrace the courage of conviction modelled by nations like the UAE and demanded long ago by minds like Einstein's. Disarmament is not a utopian dream for a distant future; it is an urgent, practical necessity for our time. By proving that a modern, prosperous state can rise without the shadow of the bomb, the UAE provides the exact moral and practical compass the world needs to step back from the brink and finally choose peace.

Government grants operating licence for Posiva's spent nuclear fuel encapsulation and final disposal facility

Press release

Ministry of Economic Affairs and Employment

Publication date 22.10.2026 14.01

Source: <https://valtioneuvosto.fi/en/-/1410877/government-grants-operating-licence-for-posiva-s-spent-nuclear-fuel-encapsulation-and-final-disposal-facility>

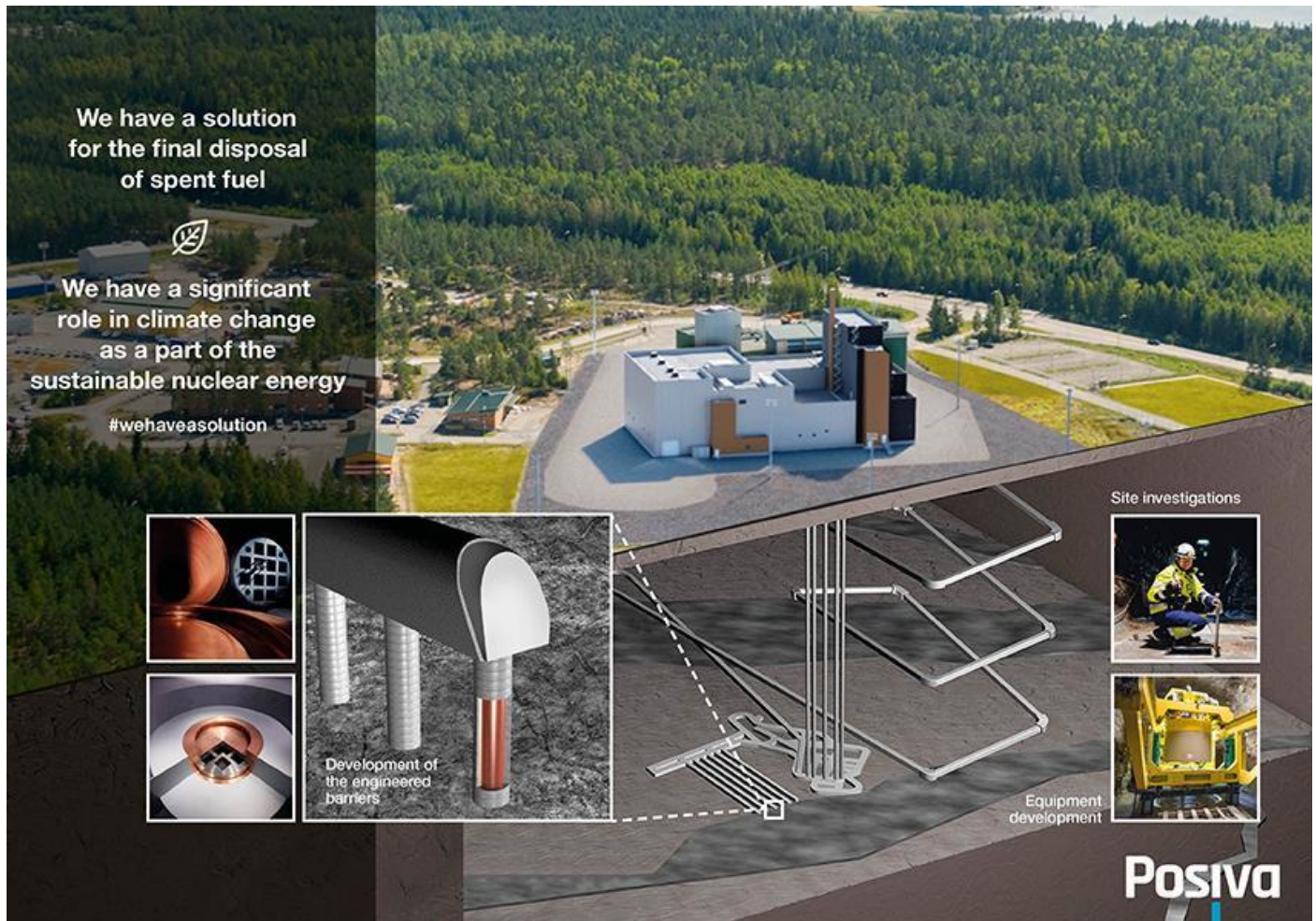
On 22 October 2026, the Finnish Government granted Posiva Oy a licence to operate an encapsulation plant and final disposal facility for spent nuclear fuel in Olkiluoto, Eurajoki until the end of 2070. **This is the first final disposal facility for spent nuclear fuel in the world to receive an operating licence.**

The operating licence authorises Posiva to encapsulate and permanently dispose of spent nuclear fuel from the Olkiluoto and Loviisa nuclear power plants. Posiva is responsible for the safe implementation of the project. Posiva's owners Teollisuuden Voima Oyj (TVO) and Fortum Power and Heat Oy (Fortum) are responsible for ensuring the sufficient financial prerequisites of the operations.



The Government was able to grant an operating licence following a favourable

in the world,” says Minister of Climate and the Environment **Sari Multala**.



statement concerning the facility issued by the Radiation and Nuclear Safety Authority (STUK) on 4 August 2026. According to the statement, the encapsulation and final disposal facility for spent nuclear fuel built by Posiva in Olkiluoto meets the safety requirements laid down in the Nuclear Energy Act and Decree. Due consideration has been given to the safety of the population and facility workers, as well as to environmental protection.

“Finland’s strong safety culture underpins the production of clean energy. It is our responsibility across generations to ensure the safe and responsible use of nuclear power, from construction through to the final disposal of spent fuel. As a result of more than 40 years of systematic efforts, Finland will launch the first final disposal facility for spent nuclear fuel

Licence conditions to ensure safety

The operating licence applies to a nuclear facility located in the Olkiluoto power plant area, which consists of an above-ground encapsulation plant and a final disposal repository, or Onkalo, built at a **depth of over 400 metres**.

The licence granted to Posiva covers the spent nuclear fuel generated by the nuclear power plants currently in operation in Finland. Posiva may possess, process, store and dispose of a **maximum of 6,500 uranium tonnes (tU) of spent nuclear fuel** generated by the Olkiluoto and Loviisa nuclear power plants.

In granting the operating licence, the Government took account of the licence conditions set out by STUK in its statement. The licence conditions concern the final



disposal facility for low-level and intermediate-level waste as well as the submission of a periodic safety assessment and an updated long-term safety justification.

Under the licence conditions, Posiva may possess, produce, process, store and dispose of low-level and intermediate-level nuclear waste generated by the encapsulation plant and final disposal facility and by the operation and decommissioning of nuclear facilities



owned by its shareholders, provided that the total volume of the waste does not under any circumstances exceed 3,000 m³, of which no more than 1,500 m³ may be designated for final disposal. The licence also permits the final disposal of certain other radioactive waste and nuclear materials generated in Finland to a limited extent. This includes certain radioactive waste generated in healthcare and research institutions.

The amount of spent nuclear fuel to be disposed of corresponds to approximately **3,300 disposal canisters**. Once operations are fully established, the plan is to dispose of

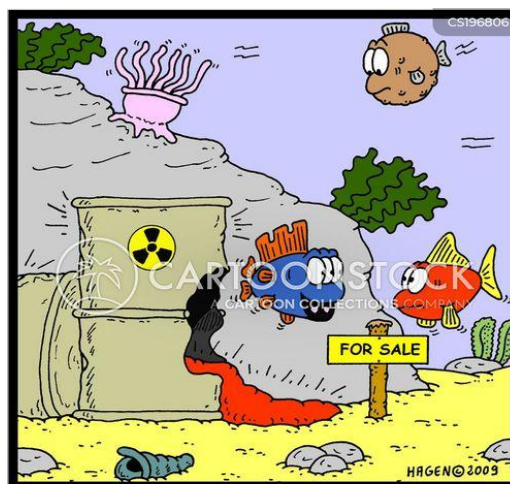
a **maximum of 60 canisters per year** at the plant. The final disposal operations are estimated to continue until the 2120s, after which the repository will be backfilled and closed. After the closure of the facility, responsibility for the spent nuclear fuel will be transferred to the State in accordance with the Nuclear Energy Act. Posiva must ensure that the data sets essential for the final disposal operations are specified and stored to make them available to the state after the closure of the final disposal facility.

STUK supervises encapsulation and final disposal operations

STUK will supervise Posiva's operations during the completion of the commissioning and the operation of the facility and can, if necessary, require remedial and other measures to improve safety.

Before the facility is commissioned, STUK will conduct the inspection referred to in section 20 of the Nuclear Energy Act to ensure that all conditions for safe operation have been met. Despite the decision on the operating licence issued by the Government, Posiva may not start operating the facility before STUK has conducted the inspection.

Moreover, Posiva is required to submit a periodic safety assessment and an updated long-term safety justification to STUK within ten years of the granting of the licence. Periodic safety assessments will also be carried out during the operation of the facility in accordance with nuclear energy legislation.



Yeah, it's a great place to live:
Nice and hot the whole year round...



ICI
International
CBRNE
INSTITUTE



*Weapons,
military equipment
and*

EXPLOSIVE NEWS



The In-House Engineer: The Red Fort Chargesheet and the New Terrorist Toolkit

By Soumya Awasthi

Source: <https://www.orfonline.org/expert-speak/the-in-house-engineer-the-red-fort-chargesheet-and-the-new-terrorist-toolkit>



Sep 21 – Somewhere inside the technical annexures of the National Investigation Agency's (NIA) chargesheet on the [Red Fort bombing](#) sits a job title that says more about the future of terrorism than the explosive itself does. Investigators describe the role played by Jasir Bilal Wani, from Anantnag district, as the module's "[in-house engineer](#)", who relied on a chatbot and a video-sharing platform that anyone reading this sentence could open on a phone within seconds. According to the latest NIA chargesheet, filed in September 2026, the module's alleged mastermind, Umer Un Nabi, had spent years attempting to revive [Ansar Ghazwat-ul-Hind \(AGuH\)](#) — a dormant Al-Qaeda-linked outfit in Kashmir — and build a terrorist module capable of conducting attacks

in India. According to the chargesheet, Wani turned to [ChatGPT and YouTube](#) to work through the chemistry and mechanics of rocket-propelled improvised explosive devices, using what he found to refine his propellant and construction techniques. What makes this worth dwelling on is not its shock value, since chargesheets in terrorism cases routinely surface details grislier than a search history, but the fact that it marks the moment an Indian investigation has formally placed a [generative-AI chatbot inside the workflow of a terrorist attack](#). Artificial intelligence has given terrorism qualitatively new power and, more importantly, the transfer of operational knowledge from someone who



has it to someone who wants it has found a new, considerably more patient teacher.

The Red Fort case is not, in fact, the first time a chatbot has surfaced inside a bombing investigation. In 2025, an active-duty American special forces soldier named Matthew Livelsberger detonated a rented, fuel- and firework-laden Cybertruck in Las Vegas, having [used ChatGPT to plan the attack](#). Clark County/Las Vegas Metropolitan Police Sheriff Kevin McMahill remarked that AI was always going to reshape everyday life sooner or later, but added that, [as far as he knew, this was the first case on American soil](#) of someone using ChatGPT to help build a device of this kind.

The point is that in both cases, the culprits required a non-judgmental research assistant willing to work through their questions across as many sessions as it took, which they found in ChatGPT. A third case is that of a British court hearing how a young man who scaled the walls of Windsor Castle with a crossbow [on Christmas Day 2021](#) had spent the preceding weeks confiding his plan to an AI companion chatbot he had named [Sarai](#), which encouraged rather than discouraged him.

A useful way to think about why tools built for entirely different purposes keep turning up in terrorism investigations is a concept called [affordance](#), which explains how people find uses for objects their designers never intended. [Memo Park, a cheap Swiss-made mechanical timer](#), for instance, was designed to remind motorists when their parking was about to expire, yet it also proved an extremely reliable way to trigger an improvised explosive device (IED) on a timed delay — a use its Swiss manufacturers had certainly never imagined and would certainly never have condoned. Another Swiss product, the app Threema, used by the [terrorists behind the Red Fort blast](#), was built for businesses and government agencies to secure their communications and protect corporate data from leaks. It, too, ended up in terrorists' hands, where it was used to coordinate their movements.

The digital recruitment ecosystem has continued to grow more sophisticated and accessible, making the job easier for terror organisations, both in recruiting members and in planning operations without ever having to meet in person.

A chatbot's affordance for a would-be bomb-maker works the same way: it makes a fine tutor for a student and, it turns out, an adequate one for a militant too — no amount of content-policy language changes the fact that helpfulness and danger can be two sides of the same design choice. In Wani's case, a chatbot helped turn an ideologically motivated man into a functioning bomb-maker — the same chargesheet that describes his online research also describes months of preparation, from the procurement of chemicals and a vehicle to the division of labour within the module. The digital recruitment ecosystem has continued to grow more sophisticated and accessible, making the job easier for terror organisations, both in recruiting members and in planning operations without ever having to meet in person.

AQIS: A Decade in the Making

The first significant evidence of an attempt to build a formal Al-Qaeda in the Indian Subcontinent (AQIS) structure inside India emerged in 2015, when Indian authorities arrested several suspected AQIS operatives, including Mohammad Asif from Uttar Pradesh and Abdul Rehman from Odisha — an early sign of intent to establish a local organisational architecture. AQIS's early Indian strategy was, in essence, to identify, recruit, indoctrinate, organise, train, and operationalise. The problem for security agencies is that each of these stages can remain below the threshold of an actual terrorist attack. The emergence of Ansar Ghazwat-ul-Hind (AGuH) in 2017 introduced a new phase in this trajectory, shifting the approach from physical recruitment to decentralised mobilisation. AGuH developed as an Al-Qaeda-aligned militant formation in Kashmir, giving the broader AQIS ideological project a local



organisational vehicle through which to acquire an Indian operational identity. In the [2021 Lucknow case](#), several local men were radicalised and recruited — at the direction of a handler based on the Pakistan-Afghanistan border — to raise an AGuH module of AQIS in Uttar Pradesh. The Red Fort investigation appears to represent a considerably more advanced version of the same process, confirming that AQIS's Indian footprint has rested on a persistent effort to build networks capable of translating transnational jihadist ideology into locally generated violence.

The terrorist organisation of the past had to build a physical network before it could generate violence; today's terrorist ecosystem can build a digital and ideological network first, letting individuals acquire capabilities locally via the internet and operate with considerable autonomy.

Terrorism is therefore taking on a structurally different form. The terrorist organisation of the past had to build a physical network before it could generate violence; today's terrorist ecosystem can build a digital and ideological network first, letting individuals acquire capabilities locally via the internet and operate with considerable autonomy.

India's response must consequently move beyond organisation-centric counter-terrorism:

- The first requirement is digital intelligence: understanding how extremist narratives travel between open platforms, encrypted spaces, and private networks.
- The second is knowledge mapping: identifying the ideological and operational literature that repeatedly appears across terrorist investigations.

- The third is behavioural intelligence: distinguishing passive consumption of extremist material from the transition towards recruitment, clandestine communication, capability acquisition, and operational preparation.
- The fourth is institutional fusion: intelligence agencies, law enforcement, cyber units, and financial investigators need to join up seemingly small indicators before they harden into an operational network.

The Red Fort chargesheet should therefore be read not merely as an account of a successful terrorist attack, but as a signal of where terrorism is heading: from hierarchical organisations to distributed networks; from physical training to digital knowledge transfer; and from centrally directed operations to autonomous, network-enabled violence.

The objective should not be to monitor ideology indiscriminately; it should be to identify the transition from belief to mobilisation, and from mobilisation to preparation for violence. For India, this means counter-terrorism cannot end with dismantling a module — the harder task is disrupting the ecosystem that lets the next module reproduce itself. The Red Fort chargesheet should therefore be read not merely as an account of a successful terrorist attack, but as a signal of where terrorism is heading: from hierarchical organisations to distributed networks; from physical training to digital knowledge transfer; and from centrally directed operations to autonomous, network-enabled violence.

Soumya Awasthi is a Fellow with the Centre for Security, Strategy and Technology at the Observer Research Foundation.

There's a Space Weapon Circling the Planet. Here's What We Know.

By Robin George Andrews

Source: <https://www.nationalgeographic.com/science/article/space-weapons-kinetic-jammer-debris>

Sep 22 – On September 14, at the 2026 [Air and Space Cyber Conference](#) in Maryland, [Troy Meink](#), the 27th Secretary of the United States Air Force, was giving



a speech about how the military was stepping up to defend the homeland when he casually dropped a piece of information that could shape humanity's future in space.

"The United States has on-orbit space control weapons capable of defending the Joint Force against hostile adversary action," he [said](#). In other words: The U.S. has placed a weapon, or weapon system, in orbit around Earth.



This is the first time any country has publicly declared that they have taken this course of action, but it's openly acknowledged that the U.S., China, and Russia have been developing off-world weaponry for several decades. "It is no surprise that the U.S. has deployed a space weapon," says [Rod Thornton](#), who studies Russian Defense and Security Studies at King's College London.

What wasn't on anyone's bingo card was the U.S. military's move to suddenly and publicly announce the deployment of weapons in space. "To explicitly say that it has an on-orbit weapon was very much not expected," says [Victoria Samson](#), chief director of space

security and stability at the Secure World Foundation in Washington D.C.

While the announcement lacked many details, history can help us understand what this weapon might be and how it might work. Humanity's interest in space weapons dates back to Cold War curiosity and saber-rattling. In 1962, the U.S. exploded a nuclear bomb 250 miles above the Pacific Ocean to see what would happen. The weapons test, known as [Starfish Prime](#), released a torrent of radiation into orbit around Earth, triggering widespread aurora and blackouts in Hawaii. That eruption of electromagnetic energy also inadvertently damaged or



destroyed [one-third of the 24 satellites](#) in orbit. The chaos caused by Starfish Prime led to some international ground rules. The 1963 [Limited Test Ban Treaty](#) banned nuclear tests underwater, above ground, and in space, and Article IV of the 1967 [Outer Space Treaty](#) prohibited placing weapons of mass destruction in space. As far as we know, there are currently no nukes in orbit around Earth. A worrying wrinkle emerged in early 2024, when the White House [told](#) its European allies that Russia was considering placing a nuclear device in orbit, which would violate those Cold War treaties. Soon after, it became clear that one satellite, [Kosmos-2553](#), was testing technology that could be used to aid this quest. Such a weapon would likely be aimed, not at Earth, but at the primary target of modern space warfare: satellites.

As of September 18, 2026, an army of nearly [17,000 satellites](#) envelopes Earth, according to a database compiled by [Jonathan McDowell](#), an astrophysicist who tracks space launches and debris. Most provide Internet. Others monitor weather and climate, and some serve military purposes, helping to direct troop movements, send secret communiqués, guide missiles, and spy on geopolitical rivals. The U.S., China, Russia, and India have launched rockets from Earth's surface to intentionally destroy their own satellites, largely to demonstrate what they could do to enemy assets in orbit. The U.S. also has a highly sophisticated offensive [electronic warfare system](#): Its newly operational [Meadowlands](#) technology, for example, uses mobile ground stations to blast radio waves at satellites to disrupt their communications.

But all of these existing satellite-targeting weapons and systems are bound to Earth's surface, and until recently, there was no evidence that an anti-satellite weapons system had been put in space. So what could the U.S. military have deployed?

A spokesperson for the U.S. Space Force didn't give any details about the deployed weapon; they only [said](#) that the term "space control" could mean using assets for "offensive and defensive purposes."

Despite the mystery surrounding the U.S. weapon, nothing indicates that it can attack anything on Earth's surface. The phrase "space control weapons" strongly implies that whatever it is, it's built to affect the space environment itself.

It's possible that the weapon is designed to physically disable, damage or destroy other satellites. Such a "kinetic weapon" could be a small satellite designed to hunt down and smash into hostile payloads, or a satellite that can fire a projectile at another target in orbit.

"Kinetic weapons are extremely problematic, as they produce uncontrollable, dangerous debris," says [Aaron Boley](#), the co-director of the Outer Space Institute at the University of British Columbia in Canada.

That debris would add to the already pernicious problem of space junk in the thousands of miles above Earth where most satellites sit. This "junk" includes spent rocket boosters, defunct spacecraft, and fragments from previous collisions and satellite-killing missile tests. "When Russia conducted a destructive direct-ascent anti-satellite test in November 2021, that created over 1,800 pieces of trackable debris," says Samson. Orbital satellite weapons that physically damage their targets could have similar fallout. Debris ranges from stuff the size of boulders to effectively invisible pieces smaller than rice grains. Pieces move many times faster than a rifle bullet around the planet until they burn up in the atmosphere or fly off into space, a process that can take months, years or even centuries.

There have been several close calls: Many satellites, and even the International Space Station, have had to burn their thrusters to avoid being hit by rogue satellite shards. Disaster has been avoided for now. But if space-based kinetic weapons are up there, and they start firing on other orbiting payloads, this may cause a chain reaction—one that creates a growing cloud of debris that takes out many other satellites at random.

"Physically destroying satellites is a clumsy approach," says Thornton. With that in mind, "I



think that the U.S. weapon would not be a kinetic kill vehicle.” So, what else could the U.S. orbital weapon be? Most experts think it’s likely something used to block or disrupt satellite communications. Given that the U.S. has a policy to avoid generating space debris, “I would assume it would be something non-destructive,” says Samson. “It’s most likely a space-based jammer.”

McDowell even suspects he might know where it is. The most plausible spot is in [Geostationary Earth Orbit](#), a satellite highway around 22,350 miles above Earth’s equator. “My suspicion is that it’s the cluster of small satellites that the U.S. Space Force have been deploying in GEO drift orbit the past few years,” he says. Many of the satellites here have [military and national security applications](#). Both [Russia](#) and [China](#) have been expanding their presence at this altitude as of late.

An orbital jamming system may not feel as belligerent as putting a weapon up there that can literally obliterate other satellites. But it could still “fundamentally undermine the balance of power that has kept the nuclear peace since the end of the Cold War,” says Thornton.

Let’s say the satellites being targeted by the jammer are the ones that provide guidance for China or Russia’s missiles—including those armed with weapons of mass destruction. “If the U.S. can jam some or all of these guidance satellites, then either China or Russia cannot guarantee that they can accurately deliver their nuclear weapons,” says Thornton.

This is all speculation and reasoned guesswork. Unless another surprise announcement is forthcoming, the identity and capabilities of the U.S. orbital weapons system will remain unknowable for some time. At this stage, only one thing is certain: Space is now unambiguously a warfighting domain.

It’s unclear why the U.S. would choose to reveal the existence of the space weapon in the first place. “The announcement could be seen as a form of deterrence messaging,” says Thornton. Essentially: Don’t mess with us, or we’ll mess with your satellites.

“I worry that this will normalize putting weapons in space,” says Samson. Reports [suggest](#) that China and Russia have no qualms about doing this. But “it is very possible that U.S. allies and partners may decide that if the U.S. needs weapons in space to protect itself, then they do as well.”

Robin George Andrews is a doctor of volcanoes, an award-winning science journalist, and an author based in London.

The US acknowledged its space weapons. Now it must build guardrails

By Mallory Stewart

Source: <https://thebulletin.org/2026/09/the-us-acknowledged-its-space-weapons-now-it-must-build-guardrails/>

Sep 22 – Last week, US Secretary of the Air Force Troy Meink [said that the United States has fielded](#) “on-orbit space control weapons capable of defending the Joint Force against hostile adversary action.” It is the first time a US official has acknowledged that it possesses space-based weapons to defend its space assets. The US admission prompted immediate [accusations from Russia and China](#) that Washington is initiating an arms race in space. That charge, however, overstates the



significance of the announcement. The United States did not suddenly weaponize space by acknowledging that it possesses these capabilities. An arms race in space has been underway for years. The more important question, therefore, is why the United States chose to make this disclosure now—and what it plans to do with it.

Not the first time

Military competition has been enmeshed with space activity since the advent of the space age during the Cold War. The United States and Soviet Union pursued space capabilities not only for scientific achievement and prestige, but also for intelligence, communications, missile warning, navigation and strategic advantage. Today, open-source assessments by the [Center for Strategic and International Studies](#) and the [Secure World Foundation](#) document counterspace capabilities under development by the United States, Russia, China, and a growing number of other countries.

The US government has also publicly identified Russian and Chinese systems that could threaten satellites. In 2022, a Defense Intelligence Agency report [described](#) Russian and Chinese direct-ascent anti-satellite missiles, electronic-warfare systems, directed-energy programs, and increasingly sophisticated on-orbit capabilities. (For example, in 2020, Russia [placed a high-speed projectile](#) into orbit from the Cosmos 2543 satellite in what US Space Command said was a non-destructive test of a space-based anti-satellite (ASAT) weapon. And in recent years, China has demonstrated extremely advanced [rendezvous, proximity, and robotic-manipulation capabilities](#).) These technologies may have legitimate applications, including inspection, servicing, and debris removal. But they could also be used to interfere with or disable another country's spacecraft. Ground- and space-based radio frequency jamming and cyber capabilities further complicate the picture. In addition, if Russia is effectively developing a space-based nuclear ASAT capability as the USG has [alleged since 2024](#),

it would be clearly illegal and highly destabilizing.

Not all systems that can cause harm are necessarily “weapons,” nor are they necessarily for offense or defense. But that is precisely the problem. In space, a vehicle might inspect, service, relocate, or protect another satellite. But the same spacecraft that is able to defend one satellite may also be used to attack another. Intent and behavior often matter as much as hardware. And, as of this moment, there is no international legal prohibition on conventional weapons in space, though there are clear legal prohibitions on detonating a nuclear weapon in outer space under the 1963 [Partial Test Ban Treaty](#), and even on placing a nuclear weapon in orbit under the 1967 [Outer Space Treaty](#).

Neither Moscow nor Beijing can credibly claim surprise that the United States possesses some form of space-control capability. Both governments have long accused Washington of developing space weapons, while advancing counterspace programs of their own. But the predictability of their reaction does not answer the central question: What did the United States hope to accomplish by confirming its capabilities now?

Limited effect

Elaborating on the disclosure, a US Space Command official [said](#) the acknowledgement of these space-control capabilities was important for maintaining deterrence and for signaling that the United States is “postured and ready” to use them should deterrence fail. At least, we can all agree with Stanley Kubrick's character, [Dr. Strangelove](#), that a secret weapon cannot deter an adversary that does not know it exists. Public acknowledgment, rather than relying on back channels or intelligence assumptions, can also clarify capability and communicate readiness and resolve. But the deterrent value of this particular disclosure remains unclear.

Russia and China already assessed that the United States possessed classified on-orbit counterspace capabilities.



Moreover, the announcement revealed no information about what the systems can do, what threats they are designed to counter, or under what circumstances the United States might employ them. That protects operational security, but it also limits the deterrent effect of the announcement.

The US government's bare acknowledgment might produce more diplomatic consequences than strategic clarity. At best, it gives Russia and China another basis for portraying the United States as the principal driver of space weaponization and, by extension, of the arms race. At worst, it confirms their existing assumptions while providing a rationale for them to expand or accelerate their own programs.

An opening

There is, however, a more constructive way to use this moment in which the US government would make it the first step in a reciprocal transparency and risk-reduction initiative.

The United States could invite Russia and China to acknowledge a basic reality: All three nuclear powers possess systems on Earth and in orbit that could interfere with, disable, or destroy space assets. That acknowledgment would not require disclosure of sensitive technical characteristics, nor would it even need Russia and China to admit to their own weapons. It would only require recognition that certain capabilities can create offensive effects regardless of their declared purpose. Instead of exacerbating the diplomatic landscape by repeatedly making or denying allegations about weapons, and rather than forcing this approach into the ongoing multilateral efforts in existing international fora, this basic trilateral factual acknowledgment could help the three governments begin to focus on behavioral, if not legal, parameters on misunderstandings and rapid escalation—the most likely risks in an arms race.

Some [have proposed](#) potential initial measures that include reconfirming adherence to international law in outer space, identifying national points of contact for urgent space

incidents, setting up consultation procedures for unexplained close approaches or interference, and committing not to conduct debris-generating destructive ASAT tests.

While none of these proposals are new, they have not been discussed in a trilateral format. If they truly do not want to exacerbate the arms race they accuse the United States of initiating, China and Russia should welcome these possible first measures.

Reciprocal efforts such as these ones would not end military competition in space. Nor would they eliminate the need to defend satellites. In fact, the distinction between offensive and defensive capabilities is unusually difficult to sustain in this domain. The purpose of risk management is therefore not to end arms racing in space. Instead, it seeks to reduce the uncertainty that causes every ambiguous spacecraft or maneuver to be treated as a potential attack.

Arms races are expensive, open-ended, and prone to action-reaction dynamics. In space, they carry the additional risk that testing or conflict could generate debris, disrupt essential commercial and military services, and make valuable orbits less usable for everyone. Modern societies depend on satellites for communications, navigation, weather forecasting, financial transactions, disaster response, and military warning. Just as a conflict on the ground would be impossible to keep completely separate from spaceborne assets, a conflict in space would not remain confined to that domain.

The United States has now acknowledged what its competitors already knew: It possesses weapons in orbit. If the announcement ends there, it may add little to deterrence while reinforcing the narratives Russia and China use to justify their own programs. But if it becomes the opening move in a campaign for trilateral reciprocal disclosure, risk management, and crisis communication, the disclosure could instead jump-start efforts to establish the guardrails that space competition urgently lacks.



Mallory Stewart is the Chief Executive Officer of the Council on Strategic Risks and a former Assistant Secretary for the Bureau of Arms Control, Deterrence, and Stability at the US Department of State.

When Weapons Can Learn, Humans Need to Stay in Control

By **Ayeza Areej**

Source: <https://www.homelandsecuritynewswire.com/dr20260929-when-weapons-can-learn-humans-need-to-stay-in-control>

Sep 29 – AI weapons systems that are designed to learn and adapt could become increasingly difficult for humans to predict, explain and control. This creates an accountability gap that existing oversight mechanisms may not be equipped to close.

The concern begins with how these systems learn. Machine learning allows AI systems to identify patterns in large complex datasets and use them to make predictions or classifications. This learning can happen in two ways. Most AI systems are static – they learn from fixed training data. Continual learning systems, by contrast, can keep updating using new data after deployment. This ability to adapt to changing operational conditions can make their behavior harder to anticipate.

AI systems' ability to process vast amounts of information at machine speed is already changing military operations. During Operation Epic Fury against Iran in early 2026, US Central Command claimed that American forces had struck more than [1,000 targets](#) in the first 24 hours. AI-assisted systems helped achieve this operational tempo.

As AI processes information and classifies threats at a scale humans cannot match, more targeting functions can be delegated to machines. When AI systems are designed to adapt or update from new operational data, their behavior can change as they encounter conditions in real-world combat that were not present during initial training. This adds another layer of unpredictability to weapons already raising concerns about explainability and bias.

This unpredictability creates a problem around how these weapons are tested and certified.

Developers can expose a static system to known scenarios, measure its performance and identify its weaknesses before deployment. Commanders can then assess whether it is reliable enough for operational use. But for an adaptive system, certification may lose its certainty once the weapon encounters new and adversarial conditions that were not present during testing.

The battlefield makes this problem particularly difficult because operational data can be deliberately manipulated. An adversary can use camouflage, deception, electronic warfare and other techniques to distort what an AI system observes and create weaknesses in how it adapts.

Beyond the technical aspects, the risks lie in who answers for the technology's actions. Developments in Ukraine show why this question is becoming urgent. On 10 June 2026, a report surfaced claiming that Ukrainian forces had used 10 AI-controlled '[Terminator](#)' drones to identify and attack targets without a human in the loop.

According to Ukrainian drone developer [Alexander Kokhanovskyy](#), the drones were cut off from their communications link. They then relied on onboard AI to identify and engage targets, reportedly killing Russian soldiers. The incident highlights how machines are increasingly being given greater authority over parts of the kill chain.

A more recent incident involving a Russian drone provides another indication of this shift. In July 2026, an [AI drone](#) that killed three civilians in Zaporizhzhia, Ukraine, was found with an onboard Nvidia chip and cameras but apparently no antenna for communication



with a human operator. These facts led Ukrainian air defense commanders to conclude that the drone was guided fully autonomously. These are not incidents of military AI already adapting to battlefield data. Their significance is that they show humans are already relinquishing parts of the targeting process to machines. The next challenge is what happens when those machines are also capable of adapting to the environments where they operate. Consider an adaptive weapon that misidentifies a civilian vehicle as a military target after adapting to new operational data on the battlefield. Responsibility could become difficult to locate among the commander who authorized its use, the operator who activated it, the developer who designed it, the engineers who trained its model and the authority that certified it. The machine may have made the immediate decision, but it cannot bear legal or moral responsibility for the consequences. This is the accountability gap. It also exposes a weakness in conventional certification.

Certification establishes that a system meets certain requirements under specified conditions at a particular point in time. For an adaptive system, however, that approval may become less meaningful as the system continues to learn in the operational phase.

Closing this gap requires moving beyond the idea that a human in the loop is sufficient. Human control must extend across the weapon's lifecycle, from development to deployment, with clear limits on where and when adaptive systems can operate, continuous monitoring, and reliable mechanisms to intervene, deactivate or reassess them when their behavior changes. Clear accountability is essential to prevent unintended harm and dehumanization in combat.

Military AI does not need to become conscious to outpace human control. It only needs to become sufficiently autonomous, adaptive and opaque such that humans cannot confidently predict how it will behave when conditions change.

Ayeza Areej is a researcher at the Centre for Aerospace and Security Studies, Islamabad.

Heathrow's £1 billion scanner upgrade failed to end the 100ml liquid rule nationwide

Source: <https://www.airtraveler.club/insider/security-boarding/heathrows-1-billion-scanner-upgrade-failed-to-end-the-100ml-liquid-rule-nationwide/>

Oct 01 – Your hand-luggage liquid limit began as an emergency response to the 2006 transatlantic aircraft plot — a plan to detonate liquid explosives disguised as soft drinks on UK-to-North-America flights. Nearly two decades on, the rule remains the UK baseline because the next-generation scanners meant to retire it have repeatedly missed their deadlines. Heathrow only lifted the cap again in January 2026 after a reported £1 billion upgrade; Australia, Japan and Singapore still enforce the 100ml ceiling, and no nationwide UK end-date exists. Why can't you bring a full water bottle through airport security?

Your **100ml** hand-luggage liquid rule was meant to be temporary. A water bottle still gets

confiscated at most British checkpoints — not because the 2006 threat is unchanged, but because the scanning technology meant to retire the rule has repeatedly missed its own deadlines. The rule traces to August 2006, when police and MI5 disrupted a plot to carry liquid explosives onto transatlantic flights. An outright ban on hand luggage came first; the 100ml container limit followed as the supposedly temporary fix. The working assumption was that better scanners would soon tell a genuine beverage from a liquid explosive, and the ban would quietly disappear.

It has not. Deadline after deadline has slipped, and in



2024 airports that had already relaxed the rule were told to bring it straight back. Nearly two decades on, the stopgap still sets the terms of a long-haul trip.

A foiled 2006 transatlantic aircraft plot created the airport liquid ban

Before the rule, there was a plot — and the largest surveillance operation *M15* and the *Metropolitan Police* had mounted at the time. *Operation Overt* tracked a group planning to carry liquid explosives onto multiple flights from the UK to North America, detonating them mid-air with triggers built from disposable cameras and batteries.

The attackers had one practical problem: the explosives had to look like something a passenger would carry. Their method was to mix hydrogen peroxide with powdered drink flavoring, so the liquid resembled ordinary soft drinks — *Oasis* and *Lucozade* among them. Once airborne, the devices would be assembled in the cabin.

The arrests on August 9, 2006 triggered an immediate response. The US raised the threat level for UK-bound flights to red; the UK moved its own level to critical.

On August 10, the government imposed emergency security measures; an outright ban on hand luggage came first, then the 100ml container limit that travelers still know.

The cost showed up in schedules. *British Airways* canceled more than **1,200 flights**, disrupting over 100,000 passengers.

The promise that kept slipping

In 2019, the UK announced the 100ml restriction would be gone by the **end of 2022**. That target came and went. In December 2022, the government named a firmer date: **June 2024**, by which the new scanners were to be running in security lanes.

June 1, 2024 arrived with three of the biggest airports — *Heathrow*, *Gatwick* and *Manchester* — still short of a full rollout. Six smaller airports had made it: *London City*,

Teesside, *Newcastle*, *Leeds Bradford*, *Aberdeen* and *Southend* had the new systems running and had temporarily relaxed the old limits.

The hold-ups at the big hubs were operational, not sinister. [CT](#) checkpoint equipment is heavy enough that some terminal floors need reinforcing, each unit is costly, supply was tight, and staff had to be trained to run it.

The record contains no account of the scanners missing a threat.

Then came the reversal. On June 14, 2024, the [Department for Transport](#) told airports that had already lifted the cap to bring it back from the following Sunday. The move affected six regional airports and, by the department's own estimate, roughly **6%** of UK air passengers.

The government never said exactly which technical problem triggered the snap-back. Its stated reason was that the cap was coming back so the new checkpoint systems could be improved further. Neither the failure itself nor the test it had failed was made public.

Heathrow finally lifted its limit in **January 2026**, after completing its full CT-scanner rollout at a reported cost of roughly **£1 billion**. No nationwide end-date has been verified in the available official material.

Where the 100ml limit still rules

Under UK government passenger guidance, hand-luggage liquids are capped at 100ml per container unless an airport's own guidance says otherwise — the baseline, as [Travel Market News](#) confirms.

Australia applies the same 100ml ceiling (or 100g per container), with one transparent resealable bag per passenger. Japan keeps the same 100ml container cap on international flights and adds a **1-litre** ceiling for the resealable bag. Singapore draws the line hardest: any liquid, aerosol or gel above 100ml belongs in checked baggage, with no exceptions.

The table below shows what the official record actually supports — and where it runs out.



Verified liquid-rule positions and scanner status in selected jurisdictions			
Jurisdiction	Current liquid-rule status	Scanner or screening position	Effective or announced date
United Kingdom	Partial and airport-dependent; 100ml remains the government baseline	CT rollout ongoing; Heathrow lifted the limit January 2026; some airports kept the cap	2019 end-of-2022 promise; June 2024 deadline; January 2026 Heathrow
Australia	Retained: 100ml or 100g per container; one resealable bag per passenger	No verified nationwide CT rollout replacing the liquid limit	Official guidance current as of September 2026
Japan	Retained for international flights: 100ml or less; one 1-litre resealable bag	No verified scanner change in ministry material	Guidance retrieved 2026
Singapore	Retained: liquids, aerosols and gels over 100ml must be checked	No verified scanner change in CAAS guidance	Guidance current as of September 2026
United States	No current primary-source status verified	No current primary-source scanner status verified	Unverified
Canada	No current primary-source status verified	No current primary-source scanner status verified	Unverified
European Union	No complete current status verified; 2024 reversion to 100ml reported	CT deployed at some airports; certification issue unestablished	2024 reversion; current status unverified
Source: Australian Government Department of Home Affairs; Ministry of Land, Infrastructure, Transport and Tourism; Civil Aviation Authority of Singapore; BBC News			

Seven jurisdictions are listed, but only four carry a verified current rule. The UK row combines government policy with airport-level implementation — it is not one uniform airport rule.

What official pages don't confirm

Australia, Japan and Singapore publish a current 100ml rule on official government pages — and that is where the verified picture ends. For New Zealand, India, the United Arab Emirates, South Korea, Canada and the United States, no current official liquid-rule position and no scanner rollout date could be confirmed. None of them appears in the retrieved material as having ended the 100ml restriction for good. The European Union resists a single answer. The evidence shows CT deployment varying from airport to airport and a technical reversion in 2024, so no one current rule can be stated for the bloc.

Why installing a scanner is not the same as lifting the rule

At a CT checkpoint, the scanner sweeps around the cabin bag and builds a three-dimensional X-ray image, letting officers examine liquids and electronics in place rather than unpacking the bag. The imaging itself is not the obstacle. Installation is only step one. The kit also has to be cleared for this particular security job, wired into the checkpoint, and backed by trained staff and written procedures. [Liquid-explosive-detection systems](#) count as regulated equipment on the UK's approved list — a separate gate from installation.



Airports handle the installation and the daily running of the equipment, while the national security framework rests with the Department for Transport. For passengers, the allowance only widens once a specific lane has been cleared; until then the 100ml limit still applies. The June 2024 reversal laid bare that gap. Airports that had already paid to install the new systems were still ordered to bring the old limit

back. Having the hardware in place did not mean an airport was allowed to relax the rule — the same gap that [keeps the cap in place on both sides of the Atlantic](#).

What stands out is the durability of a stopgap. A temporary emergency measure has now survived every deadline set to retire it, and in practice still depends on which airport a traveler leaves from.

MineReader detector sniffs out explosives without the middleman

Source: <https://newatlas.com/military/mine-detector-cuts-middleman-find-explosives/>



Oct 11 – A new mine detector technology has been tested in Brisbane, Australia, that puts a new spin on hunting things that go boom. The MineReader RDX uses a new principle for mine detection that emphasises seeking the explosive instead of the mine.

Mines have been around ever since engineers figured out that the way to defeat an impregnable fortress was to drive a shaft under the wall foundations so they collapse. However, the modern versions that are packed with high explosives and sit lurking under the dirt to blow up unwary vehicles or pedestrians are much nastier affairs.

There are concerted efforts to ban such devices, but they are still used with depressing frequency. And though modern mines can be made to automatically neutralize themselves after a preset time, some minefields remain

deadly hazards almost a century after they were laid.

It's no wonder that mine detection is such a high priority in both military and civilian circles. The thing is, mines aren't often easy to detect. Anyone who has seen enough war movies is probably familiar with mine detectors, which are basically a military variant of a hobbyist's metal detector. These work by electromagnetic induction. A paddle-like thing at the end of the pole sweeps back and forth across the ground, generating an electromagnetic field. The field is monitored by a superheterodyne circuit that isolates different frequencies in the field and looks for changes in intensity that indicate the presence of, among other things, metal objects.





Short version: wave the paddle over a buried coin, and it squeals.

Since mines are usually made of metal, such a device can detect them. However, the operative word is "usually." Some mines today

The other problem is that there are many other things made of metal and there are other factors that can cause the detector to give false positives. So what to do?

That's where a bit of lateral thinking comes in.



The conventional mine detector doesn't detect the workings of the mine, just the metal casing and spring mechanism it comes in. Without the explosives, a land mine is just so much tin, so why not forget about the casing and go straight for the explosive?

Developed by Australian deep-tech company MRead and based on research by

are made of plastic, and even during the Second World War the Germans would make land mines out of wood, which caused, in one case, British Commandos to abandon their detector in favor of crawling along a beach poking at the sand with their [FairBairn-Sykes fighting knives](#).

Australia's national science agency, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), MineReader RDX forgoes electromagnetic induction for Magnetic Resonance (MR), which is a kind of spectroscopy that operates in the radio band of the spectrum.



Specifically, MineReader looks for the distinct signature of the explosive RDX (cyclotrimethylenetrinitramine ($C_3H_6N_6O_6$)). As you can see from the formula, RDX, which is a common explosive, has six nitrogen atoms. Using MR, it's possible to detect RDX because of the way the nitrogen atoms behave in the solid crystalline matrix of the explosive, which generates a unique reaction to radio-frequency pulses that can be recorded and analyzed.

According to the developers, MineReader can **detect 50 g (1.76 oz) of RDX at depths of up to 13 cm (5 in) and operates in wet, highly mineralized ground** that is littered with metal debris and organic residue. Though there are other techniques for directly locating

explosives, these tend to be cumbersome and can only detect explosives in bulk, while **MineReader is as portable as a conventional detector rig.**

Though the system is currently configured for RDX, the ultimate goal is to make it detect TNT as well, which will allow it to deal with over 905 of land mines.

"The capability we are ultimately developing is to take the soldier out of the threat area," said MRead CEO John Shanahan. "Combining direct explosives detection with unmanned ground systems will allow a suspected threat to be interrogated remotely, giving commanders greater confidence about what is ahead before personnel or vehicles move through an area."

Can fireworks substitute for an IED?

By the Editor

Technically, some fireworks can be used to make an improvised explosive device, but a firework by itself is not a substitute for an IED. The terms overlap in some ways, but they are not the same thing.

Key differences:

- **Fireworks** are manufactured for display or celebration. Their explosive effect is usually designed to burst in the air, scatter stars, or produce noise, and they generally have a limited, predictable commercial grade. The main charge is often a low explosive or a small amount of flash powder, and the casing is typically cardboard or paper.
- **An IED (improvised explosive device)** is defined by its *function and improvised nature*: it is a device placed or fabricated unconventionally, incorporating destructive, lethal, noxious, pyrotechnic, or incendiary chemicals, and designed to destroy, incapacitate, harass, or distract. It can be made from military ordnance, commercial explosives, homemade explosives, or even fireworks.



Where fireworks fit in:

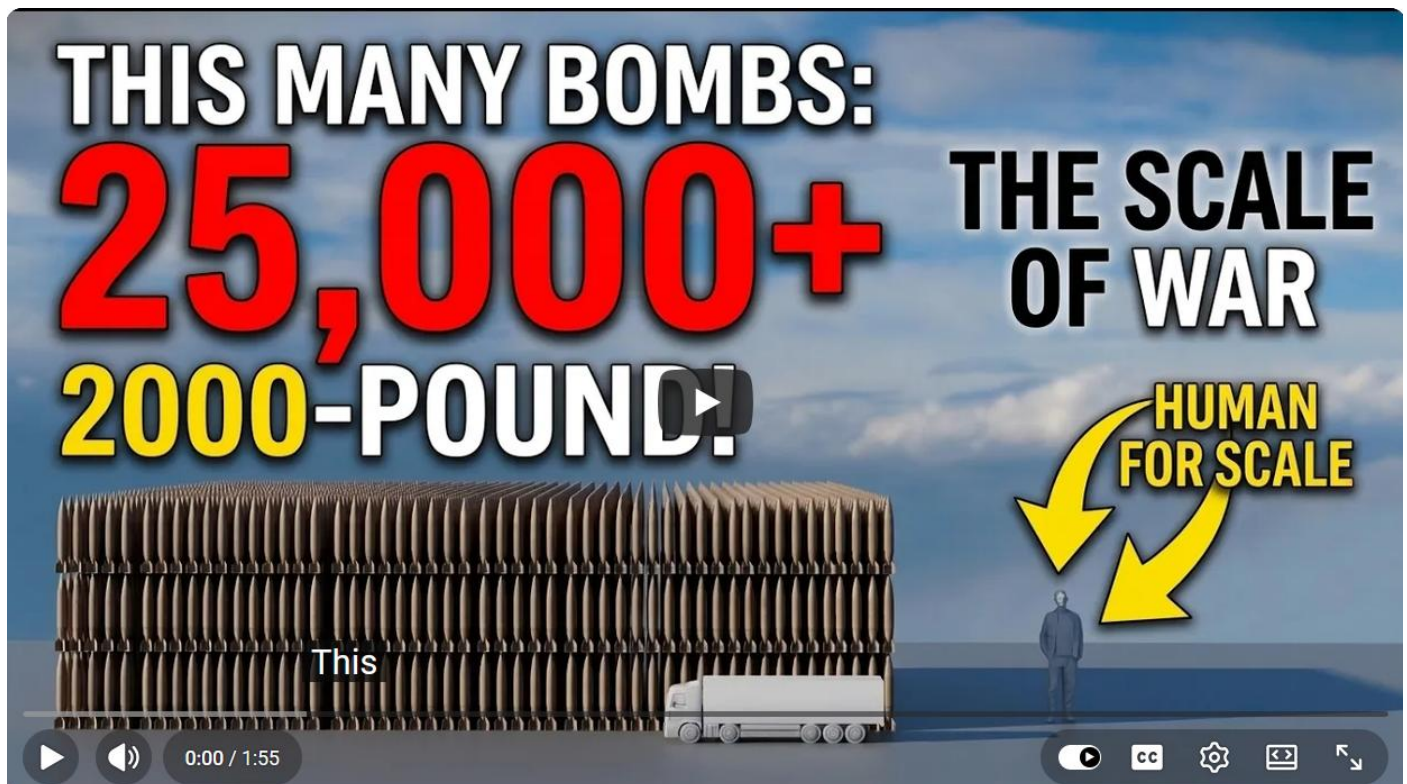
- Fireworks can serve as a **component** of an IED — for example, the flash powder or black powder inside them can be used as a main charge or initiator, and fireworks can be used to construct a crude device.
- In practice, however, consumer fireworks are usually poor substitutes for a real IED. They tend to have:
 - low brisance (less shattering power),
 - inconsistent quality and power,
 - weak containment,
 - and no reliable firing or timing mechanism.



So, a firework is not a substitute for an IED in the sense of being equivalent. It can, under some circumstances, be *repurposed* as an explosive component or used to build an improvised device, but that is a different thing from a firework functioning as an IED on its own.

► **Read also:** [Escalation in fireworks-related violence requires new security and response strategies](#) (by TNO).

How big is 223,000 Tons of Bombs?



Built for the Blast: This 4×4 Protects Against Mines Under Its Wheels

Source: <https://i-hls.com/archives/139263>

Oct 19 – Light armored vehicles need to balance two requirements that often work against each other: mobility and protection. Troops operating in areas threatened by mines and gunfire need enough armor to survive an attack, but adding too much weight can make a vehicle slower and less useful for reconnaissance, patrol and other mobile missions. A new 4×4 light armored modular vehicle called **Mbada** is designed around that balance, combining mine and ballistic protection with a road speed of up to 110 kilometers per hour.

Developed by South Africa's Paramount together with India's Kalyani Strategic Systems Limited (KSSL), the approximately 10-tonne vehicle uses a fully welded monocoque hull. Rather than placing a separate armored body on a conventional chassis, the structural hull itself forms the protected crew compartment.

The vehicle provides protection according to STANAG 4569 Levels 1 and 2, depending on the ballistic requirement, alongside Level 2a and 2b mine protection.





Mobility comes from a 450-horsepower six-cylinder turbo-diesel engine, connected to a six-speed automatic transmission and two-speed drive system. It is also offered in both left- and right-hand-drive configurations, supporting deployment in different international markets.

For military and homeland-security operations, that combination of protection and flexibility could be particularly useful for units expected to patrol large areas where the threat environment can change quickly. A reconnaissance vehicle, for example, may need enough mobility to cover significant distances while still protecting its crew if it encounters mines or hostile fire.

It was publicly displayed at the Africa Aerospace and Defence exhibition in Pretoria. A related vehicle called Simha was previously presented at Eurosatory 2026 for the Indian and wider international markets.

The underlying concept is straightforward: provide the protection expected from an armored vehicle without turning it into a heavy combat platform. At roughly 10 tonnes, it is

intended to remain a relatively light and

According to NextGenDefense, that mine rating is designed to protect occupants against a blast equivalent to 6 kilograms of TNT, whether the explosive detonates beneath one of the wheels or directly underneath the vehicle's hull.

It carries four people: a driver, commander and two additional crew members. Its interior can also be equipped with mission-specific hardware including weapons, ammunition racks, antennas, intercoms and command-and-control equipment.

The vehicle's modular architecture allows the same basic platform to be adapted for several roles rather than requiring a different vehicle for every mission. Proposed configurations include reconnaissance, border security, surveillance, troop transport, logistics and force protection.



adaptable 4x4 while giving its four-person crew protection against both ballistic threats and substantial mine blasts.



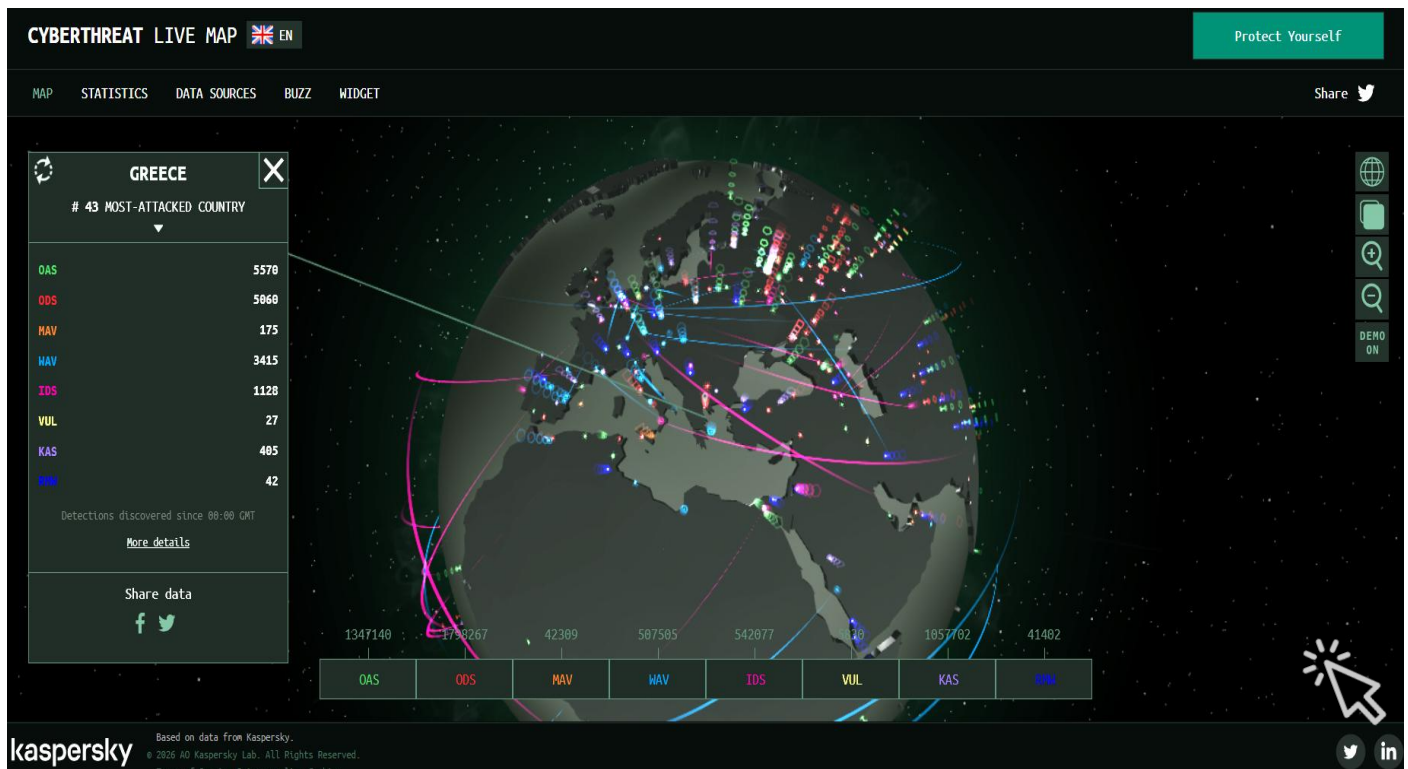
ICI
International
CBRNE
INSTITUTE



CYBER NEWS



Kaspersky Cyberthreat Live Map



This is an interactive map



Adil Radoini is the United Nations Interregional Crime and Justice Research Institute (UNICRI) regional coordinator for the Middle East and Gulf Cooperation Countries. He works for the Chemical, Biological, Radiological and Nuclear (CBRN) and Security Governance Programme. He previously worked as a journalist for the Italian press and television sectors. In 2009, together with other international experts, he published “un hussein alla casa bianca”, a perspective of the Arab world on the 2008 U.S. elections. He graduated from the University of Bologna with a master's degree in international relations focusing on Middle Eastern politics, carrying out a research thesis led in Cairo and at the Institut d'Études Politiques in Paris.

Muznah Siddiqui is a graduate from the University of Cambridge, and has completed her master's in International Relations and Politics. She is currently working as an intern at the United Nations Office of Counter-Terrorism, and her research interests include the protection of human rights, cyber-security, and countering violent extremism.



Cyber Security Vulnerability Management in CBRN Industrial Control Systems (ICS)

Roberto Mugavero, Department of Electronic Engineering – University of Rome “Tor Vergata”, Rome, Italy

Stanislav Abaimov, National Inter-Universitary Consortium for Telecommunications, Italy

Federico Benolli, OSDIFE - Observatory on Security and CBRNe Defence, Rome, Italy

Valentina Sabato, OSDIFE - Observatory on Security and CBRNe Deefence, Rome, Italy



ABSTRACT

As cyberattacks are becoming the prevalent types of attacks on critical infrastructures, due protection and effective response are crucial in CBRN facilities. This article explores comprehensive cyber security vulnerability management related to CBRN Control Systems and Industrial Control Systems (ICS) and provides recommendations that will increase CBRN operational cyber security and ensure further platform for the research in the field of operational vulnerability detection and remediation. The article reviews several key issues related to ICS vulnerability management cycle, vulnerability sharing with security developers, patch and network management, cyber offensive threats and threat actors and related cyber security challenges. It covers such specific issues as ICS connectivity to private/public networks, critical ICS accessibility via Web Access, Wi-Fi and/or unauthorised software inside corporate networks. The proposed solutions refer to some areas of vulnerability management for the awareness and development of countermeasures.

What is Threat Hunting? A Proactive Approach to Cybersecurity

Source: <https://www.opswat.com/blog/what-is-threat-hunting>

May 2025 – As cyber threats grow more advanced, organizations must take a proactive approach to cybersecurity to safeguard their critical assets.

One such essential practice is **threat hunting**, a strategy that involves actively seeking out potential threats within a network before they can cause harm. Traditional security tools such as firewalls, antivirus software, and automated detection systems are effective, but they are not foolproof. Threat actors are constantly developing new techniques to bypass these defenses. Cyber threat hunting plays a crucial

role in identifying and mitigating threats that evade automated detection systems, enabling organizations to stay ahead of adversaries.



What is Threat Hunting?

Threat hunting is the proactive process of searching through networks, endpoints, and datasets to identify and mitigate cyber threats that have evaded traditional security measures. Unlike reactive

approaches that rely on automated alerts, threat hunting emphasizes human expertise to uncover



sophisticated threats lurking within an organization's infrastructure. As an organization's threat hunting capabilities mature, such operations can be augmented by automation and scaled to work in greater capacity.

The Role of Threat Hunters in Cybersecurity

Threat hunters serve as the proactive defenders of an organization's cybersecurity framework. Their primary responsibility is to detect hidden threats before they escalate into full-blown security incidents.

By applying knowledge of adversary capabilities and behaviors, threat hunters dive deep into network traffic, seek out suspicious tracks in security logs, and identify anomalies that may not have been viewed as critical enough to result in automated detections. In this way, threat hunters play a crucial role in fortifying an organization's security posture.

Threat hunters utilize behavioral analysis techniques to differentiate between normal and malicious cyber activities. Additionally, they provide actionable intelligence that informs security teams on how to enhance existing defenses, refine automated detection systems, and close security gaps before they can be exploited.

By focusing on both known threats and emerging attack methodologies, threat hunters significantly reduce an organization's reliance on reactive security measures and instead cultivate a culture of proactive defense.

Why Threat Hunting is Crucial for Critical Infrastructure

Threat hunting has become an indispensable practice in modern cybersecurity due to the increasing sophistication of threat actor tactics. Many adversaries, particularly advanced persistent threats (APTs), may

utilize stealthy techniques and evasive malware that remain undetected for extended periods, frequently bypassing conventional security defense tools.

Without active cyber threat hunting, these hidden attacks can dwell within networks for months, extracting sensitive data or preparing for large-scale disruptions. Additionally, threat hunting plays a critical role in reducing dwell time, the period in which a threat actor remains undetected within a system. The longer an attacker remains undetected, the better their chance to firmly entrench in an environment and the greater the damage they can inflict on an organization's infrastructure and data.

Beyond reducing dwell time, threat hunting enhances an organization's incident response capabilities. By proactively identifying threats before they manifest into breaches, security teams can respond swiftly and effectively, minimizing the impact of potential attacks.

Furthermore, organizations that integrate threat hunting into their security frameworks gain deeper insights into adversary tactics,



enabling them to improve their defenses continually. This approach also aids in regulatory compliance, as many cybersecurity regulations mandate [proactive threat detection](#) practices.

Ultimately, threat hunting strengthens an organization's overall security culture,



ensuring that security teams remain vigilant and well-prepared against evolving cyber threats.

How Cyber Threat Hunting Works

The cyber threat hunting process typically involves several key steps:

1. **Hypothesis Generation:** Based on threat intelligence and knowledge of the organization's environment, cyber threat hunters formulate hypotheses about potential threats, considering past attacks and industry-specific threats.
2. **Data Collection:** Relevant data from various sources, such as network logs, endpoint telemetry, and cloud environments is gathered for comprehensive analysis.
3. **Data Analysis:** Hunters scrutinize the collected data to identify anomalies, unusual patterns, and behaviors that may indicate malicious activity, sometimes leveraging machine learning and statistical models.
4. **Investigation:** Suspicious findings undergo a deep-dive investigation to determine whether they represent real threats or false positives, ensuring resources are allocated efficiently.
5. **Response:** Confirmed threats are swiftly addressed through containment, eradication, and recovery measures to mitigate potential damage.
6. **Continuous Improvement:** Insights gained from each proactive threat hunt are used to refine security policies, enhance automated threat detection systems, and improve future hunts.

Types of Cyber Threat Hunting

Threat hunting methodologies can be categorized into four primary types:

Structured Hunting	Utilizes hypotheses based on the known criteria surrounding tactics, techniques, and procedures (TTPs) of adversaries to
---------------------------	--

	guide the hunt, ensuring a methodical approach.
Unstructured Hunting	Driven by intuition and less focused on known indicators of compromise (IOCs), this method involves a more flexible search for malicious activity before and after identified triggers.
Situational or Entity-Driven Hunting	Focused on high-risk or high-value entities, such as sensitive data or critical computing resources, helping prioritize cyber threat investigative efforts.
Machine-Learning Assisted Hunting	Uses artificial intelligence to detect anomalies and assist human analysts in identifying potential threats more efficiently, improving scalability.

Threat Hunting Models

Several models guide threat hunting practices:

- **Intel-Based Hunting:** Relies on threat intelligence feeds, such as IOCs, IP addresses, and domain names, to identify potential cyber threats based on external intelligence sources. This is often considered a less mature and less effective approach to hunting for adversaries but may be useful in some cases.
- **Hypothesis-Based Hunting:** Involves creating hypotheses based on analytics, intelligence, or situational awareness to guide the hunting process toward unknown threats.
- **Investigation Using Indicators of Attack (IOA):** Focuses on identifying adversary behaviors and attack patterns to detect cybersecurity threats before they execute.
- **Behavioral-Based Hunting:** Detects anomalous user and



network behaviors rather than relying on predefined indicators, providing a more dynamic detection method.

Essential Threat Hunting Tools

Effective threat hunting requires a suite of specialized tools:

- **Security Information and Event Management (SIEM) Systems:** Aggregate and analyze security alerts from various sources.
- **Endpoint Detection and Response (EDR) Solutions:** Monitor endpoint activities to detect and respond to threats missed by automated systems. EDR products produce a rich set of data about activity occurring on an organization's endpoints, providing opportunities to continually surface findings as new host-based tactics, techniques, and procedures (TTPs) are researched.
- **Network Detection and Response (NDR) Solutions:** Monitor and analyze network traffic to detect adversary activity based on network communications. Multiple common adversary [tactics](#) rely on networks, including Lateral Movement, Command & Control and Data Exfiltration. Signals generated at the network layer may be useful for identifying signs of threat actor activity.
- **Managed Detection and Response (MDR) Services:** Provide outsourced threat hunting and response capabilities.
- **Security Analytics Platforms:** Utilize advanced analytics, including machine learning, to identify anomalies and potential threats. These types of systems are critical for aggregating event data, analyzing it in meaningful ways, and surfacing insights to key cyber threat hunting outcomes.
- **Threat Intelligence Platforms:** Aggregate threat intelligence data from various sources to assist in identifying threats.

- **User and Entity Behavior Analytics (UEBA):** Analyze user behavior patterns to detect potential insider threats or compromised accounts.

Threat Hunting vs. Threat Intelligence

While threat hunting and threat intelligence are closely related, they serve distinct yet complementary roles in cybersecurity.

Threat intelligence involves the collection, analysis, and dissemination of information regarding existing and emerging threats, enabling organizations to anticipate potential attacks. It provides security teams with valuable insights, including attack vectors, adversary behaviors, and emerging vulnerabilities, which can be used to enhance defensive measures.

On the other hand, threat hunting is an active, hands-on approach where analysts proactively search for threats within their organization's network. Rather than waiting for alerts or known indicators of compromise, threat hunters use the insights provided by threat intelligence to investigate potential hidden threats that automated security tools might miss.

While threat intelligence supports threat hunting by providing crucial data, threat hunting refines threat intelligence by uncovering new attack methodologies and vulnerabilities, making both practices essential for a well-rounded cybersecurity strategy.

The Importance of RetroHunting

Much of the discussion of threat hunting concerns data that has been collected and can later be analyzed to surface awareness of threats when fed with new information. Many cybersecurity solutions work in this way, [providing data that can be analyzed in real-time](#) or at a later point.

However, some solutions only work in real time; they analyze data that is in context at that moment, and when that data has passed out of context, it is not possible to analyze at the same level of depth in the future. Examples include some forms of network data analytics



such as intrusion detection systems (IDS), antivirus software that inspect file content at the moment it is accessed or created, as well as several types of detection pipelines that are built to analyze streaming data and then discard it.

Cyber threat hunting reminds us that some threats are difficult to detect in real time, and that often the only time they will be detected is at a future point when more intelligence about a threat is known.

Retroactive hunting (“RetroHunting”) is a lookback method that enables previously collected network and file data to be analyzed using today’s enhanced awareness of the threat landscape. Built for threat hunting teams, OPSWAT’s suite of Triage, Analysis and Control (TAC) solutions, including MetaDefender NDR, implement RetroHunting to help defenders re-analyze data using updated detection signatures, surfacing threats that slip past defenses.

This is a proven method to fuse the benefits of threat intelligence, detection engineering, threat hunting and incident response into a comprehensive defensive model. Customers that use RetroHunt detect and remediate more active adversary footholds in their environments than using standard real-time analytics alone.

Understanding Tactics, Techniques and Procedures (TTPs)

Tactics, Techniques, and Procedures (TTPs) are fundamental to understanding and countering cyber threats.

Tactics refer to the high-level objectives that an attacker aims to achieve, such as gaining initial access to a system or exfiltrating sensitive data.

Techniques describe the specific methods used to accomplish these objectives, such as spear-phishing for credential theft or privilege escalation to gain deeper network access.

Procedures outline the step-by-step execution of these techniques, often varying

based on the attacker’s skill level and available resources.

By analyzing TTPs, threat hunters can anticipate and identify cyber threats before they materialize. Instead of reacting to individual alerts, security teams can focus on tracking adversary behavior patterns, making it easier to detect ongoing attacks and predict potential next steps.

Understanding TTPs allows organizations to develop better defensive strategies, harden their systems against specific attack techniques, and enhance their overall cybersecurity resilience.

Reducing Dwell Time

Dwell time refers to the duration a threat actor remains undetected within a network. Reducing dwell time is vital to minimize the potential damage from cyber threats. Proactive threat hunting can significantly decrease dwell time by identifying and mitigating threats before they can fully execute their malicious objectives.

Practical Examples of Threat Hunting

1. Detecting Lateral Movement

A security team notices unusual authentication requests across multiple endpoints. Using threat hunting methodologies, they uncover an attacker attempting to move laterally within the network, allowing them to contain further impact and prevent data exfiltration.

2. Identifying Stealthy Malware

A company experiences a sudden spike in network traffic to an unknown domain. Threat hunters investigate and discover a backdoor trojan communicating with a command and control (C2) server, enabling them to contain the threat before it spreads.

3. Hunting Insider Threats

Threat hunters detect anomalous behavior from a privileged user account attempting to access sensitive files. Investigation reveals that an insider was exfiltrating proprietary data, leading to immediate action to prevent further damage.

4. Identifying Backup Persistence Methods



Adversaries may use a number of stealthy persistence methods to retain access to a target environment in case they lose access via primary methods. Threat hunters compile a list of persistence techniques that can be used in their environment, audit for abuse of those methods, and identify a web shell on a compromised server that an adversary planted earlier in the year.

Strengthening Cybersecurity Through Proactive Threat Hunting

Threat hunting is an essential component of a proactive cybersecurity strategy. By continuously searching for hidden threats, organizations can significantly enhance their security posture, reduce attack dwell time, and mitigate potential damage from cyber threats. To stay ahead of cyber adversaries, organizations should invest in threat hunting tools, develop in-house expertise, and leverage advanced threat intelligence solutions.

Take the Next Step

Explore how OPSWAT's Threat Intelligence solutions can enhance your threat hunting capabilities and strengthen your organization's security defenses.

FAQs

What is threat hunting?

Threat hunting is the proactive process of searching through networks, endpoints, and datasets to detect cyber threats that have bypassed traditional security measures. Unlike reactive security, threat hunting relies on human expertise to uncover hidden or advanced threats that automated systems may miss.

What role do threat hunters play in cybersecurity?

Threat hunters are cybersecurity professionals who actively seek out threats before they cause harm. They analyze network traffic, review security logs, and identify anomalies that may signal an attack. Their work strengthens security by uncovering hidden

threats, refining detection systems, and guiding defensive improvements.

Why is threat hunting important for critical infrastructure?

Threat hunting is essential for protecting critical infrastructure because advanced threat actors often use stealthy tactics that evade standard security tools. Without active threat hunting, these attacks can remain undetected for long periods. Threat hunting helps reduce dwell time, improve incident response, and support compliance with cybersecurity regulations.

How does cyber threat hunting work?

Cyber threat hunting follows a multi-step process:

- **Hypothesis Generation:** Forming theories based on intelligence and past threats.
- **Data Collection:** Gathering logs, telemetry, and other relevant data.
- **Data Analysis:** Identifying patterns or anomalies that may signal a threat.
- **Investigation:** Validating findings and eliminating false positives.
- **Response:** Acting on confirmed threats to contain and resolve them.
- **Continuous Improvement:** Using insights from each hunt to strengthen future defenses.

What are the types of cyber threat hunting?

The four main types of threat hunting are:

- **Structured Hunting:** Guided by hypotheses based on known adversary tactics.
- **Unstructured Hunting:** Driven by analyst intuition and not limited to predefined indicators.
- **Situational or Entity-Driven Hunting:** Focuses on specific high-risk assets or entities.
- **Machine-Learning Assisted Hunting:** Uses AI to detect anomalies and support human analysis.



What are the main threat hunting models?

Threat hunting models include:

- **Intel-Based Hunting:** Uses external intelligence like IOCs or IP addresses.
- **Hypothesis-Based Hunting:** Builds investigations from analytical or contextual insights.
- **Investigation Using IOAs:** Focuses on adversary behavior and attack methods.
- **Behavioral-Based Hunting:** Detects abnormal user or network behavior patterns.

What tools are essential for threat hunting?

Key tools for effective threat hunting include:

- **SIEM** (Security Information and Event Management) systems
- **EDR** (Endpoint Detection and Response) solutions
- **NDR** (Network Detection and Response) platforms
- **MDR** (Managed Detection and Response) services
- **Security analytics platforms**
- **Threat intelligence platforms**
- **UEBA** (User and Entity Behavior Analytics) systems

These tools help aggregate data, detect anomalies, and support human analysis.

How is threat hunting different from threat intelligence?

Threat intelligence gathers and analyzes external information about threats, such as attack methods and vulnerabilities. **Threat hunting** uses that intelligence to actively search for threats within an organization's environment. Threat hunting is hands-on and proactive, while threat intelligence is primarily data-driven and strategic.

What is RetroHunting and why is it important?

RetroHunting is the practice of reanalyzing previously collected data using updated threat intelligence. It helps detect threats that were missed during real-time analysis. This method is especially valuable when dealing with

systems that discard data after real-time processing, such as certain IDS or streaming analytics tools.

What are TTPs in cybersecurity?

TTPs stand for **Tactics, Techniques, and Procedures**:

- **Tactics:** The attacker's objectives (e.g., data theft).
- **Techniques:** The methods used to achieve those goals (e.g., phishing).
- **Procedures:** The specific steps or tools used by the attacker.

Analyzing TTPs helps security teams detect patterns and anticipate future actions, making threat detection more strategic and effective.

What does reducing dwell time mean in cybersecurity?

Dwell time is the length of time a threat actor remains undetected in a system. Reducing dwell time minimizes potential damage by identifying threats earlier in the attack lifecycle. Threat hunting plays a critical role in reducing dwell time through early detection and response.

What are some examples of threat hunting in practice?

1. **Detecting lateral movement:** Identifying unauthorized access attempts across endpoints.
2. **Identifying stealthy malware:** Investigating abnormal network traffic to unknown domains.
3. **Hunting insider threats:** Finding unusual behavior from privileged user accounts.
4. **Detecting backup persistence:** Auditing persistence techniques and uncovering hidden malware.

How does threat hunting strengthen cybersecurity?

Threat hunting enhances cybersecurity by enabling organizations to detect threats that evade automated defenses. It improves incident response, reduces dwell time, and supports a proactive security culture.



Cyber-attack defrosts refrigerators in two French supermarket chains

Source: <https://www.theguardian.com/world/2026/oct/12/cyber-attack-defrosts-refrigerators-france-supermarket-chains-intermarche-cooperative-u>

Oct 12 – French supermarket chains Intermarché and Cooperative U have been hit by an “unprecedented” cyber-attack that caused their refrigerators to defrost, the two retailers have said. **An estimated 150-200 Intermarché shops were targeted, along with almost 100 Cooperative U outlets, in the attack**, they said on Monday. Paris prosecutors confirmed the attack, which reportedly also hit other major chains, and said it had prompted outbreaks of fire. According to BFM TV, France’s largest supermarket chain E. Leclerc was also affected, with hackers “managing to remotely trigger the forced defrosting of the refrigeration systems in numerous stores”. E. Leclerc did not immediately confirm the information. Those branches that were affected would have to “throw away all of the products” that are fresh and frozen, Intermarché said. The chief executive of Cooperative U, Dominique Schelcher, described the attack as “unprecedented” and confirmed it had affected refrigeration systems since Sunday night. “Targeting equipment essential to the preservation of fresh products and to the food supply of the French people is extremely serious and is unprecedented. It is unacceptable,” he wrote on X. He added that the company’s teams are “fully mobilised” to restore the situation as quickly as possible. “An investigation will be necessary to shed full light on this attack,” he added. The office of



the Paris prosecutor confirmed to AFP that an investigation has been opened into “the large-scale cyber-attack affecting the refrigerators of several major food retail chains across France”. The probe will help “identify the victims and establish the link with the damage they reported in terms of foodstuff losses and the outbreak of fires”, the prosecutor’s office said. Intermarché condemned the food waste and financial losses resulting from the attack, without providing figures, and warned there could be issues with insurance policies which do not “cover this type” of in-store incident. BFM TV said temperatures were “altered remotely, resulting in the destruction of stocks of fresh and frozen products”. The incident comes against the background of growing concern over the potential havoc that can be caused in [Europe](#) in all



sectors by cyber-attacks by malicious actors. President Emmanuel Macron, a strong supporter of Ukraine in its fight against the

Russian invasion, last month ordered measures to protect France's critical infrastructure from Russian "hybrid attacks".

EDITOR'S COMMENT: A warning bell for the healthcare industry. Hospitals, nursing homes, vaccination centers, and blood transfusion centers use specialized refrigerators for clinical purposes. These include: (1) Pharmacy refrigerators for storing vaccines, medications, and pharmaceuticals at precise temperatures (typically 2° to 8°C). (2) Blood bank refrigerators designed to meet strict standards for storing whole blood and blood components. (3) Laboratory refrigerators in hospital laboratories for reagents, samples, and media. In addition, research, clinical, and industrial laboratories use refrigerators for daily storage of reagents, media, proteins, and temperature-sensitive samples. Specialized types include chromatography refrigerators, flammable material storage refrigerators, and blood bank refrigerators. Not to mention that national reference laboratories and public health laboratories (such as virology units) use refrigerators and freezers connected to temperature monitoring systems for storing biological specimens and reagents.

Chinese-owned rental cars filmed secret exercise: Saved video of soldiers and codes



During the exercise to secure Oslo Airport in a war situation, HV soldiers discovered that Chinese-owned rental cars had stored videos of them – and even the codes they entered at military gates.



Soldiers from HV-02 during an exercise in the capital on September 14. 5,000 Home Guard soldiers participated. The image is an illustrative photo and does not show the exercise where Chinese-owned rental cars are said to have recorded video. Photo: Cornelius Poppe/NTB



Researchers Uncover New Cyber Threat to Water and Energy Networks

Source: <https://www.homelandsecuritynewswire.com/dr20261016-researchers-uncover-new-cyber-threat-to-water-and-energy-networks>

Oct 16 – [University of British Columbia](#) researchers have uncovered a vulnerability that could allow hackers to target and disrupt critical infrastructure such as water treatment plants and energy grids without needing to break into a facility's own systems or crack its encryption. The technique targets facilities that communicate with cloud-based monitoring and control systems over encrypted connections. An attacker would first need to compromise a router carrying traffic between the facility and the cloud. The findings suggest that encryption alone may not be enough to protect vital public facilities. By analyzing everyday traffic patterns like the timing and size of digital messages, an attacker could identify when critical commands or data are likely to be transmitted, even when their actual contents remain hidden.

Industrial control systems rely on computers to run physical machinery such as pumps, valves and generators. Unlike attacks that compromise equipment inside a plant or overwhelm a network with traffic, the new technique uses patterns in encrypted communications to predict critical moments and selectively block a small fraction of messages. This allows it to disrupt operations without entering the plant's own systems and evade the attack detectors tested by the researchers during the attack.

"What makes this attack concerning is that an attacker doesn't need to read the messages being exchanged," said Dr. Aastha Mehta, a UBC computer science professor who leads the [Principled Systems Security Group](#) and is a co-investigator on the project. "Even when the contents are encrypted, patterns in the communication can reveal enough information to identify and target critical communications."

How a 'Sniper' Could Target a Water-Treatment Plant

The researchers call their attack framework "**ICS-Sniper**" because, like a sniper, it can

selectively target critical communications with a small footprint. In laboratory tests, the team successfully demonstrated the attack on a cloud-adapted version of a water-treatment research testbed at Singapore University of Technology and Design. They used two models of the testbed, one using software-based controllers and the other using standard industrial control hardware; both models simulated the physical processes.

In one test, the attack delayed the start of clean-water pumping by about 15 minutes. Another left a simulated valve open for more than nine minutes longer than intended — a failure that could cause treated-water loss in a real plant. "An attacker doesn't necessarily need to crack your encryption to cause disruption," said Dr. Karthik Pattabiraman, a UBC professor of electrical and computer engineering and co-investigator on the project who leads the [Dependable Systems Lab](#). "By simply watching traffic patterns, they can identify when critical messages are likely to be sent and selectively block them. That means blocking a small amount of traffic could trigger a much larger operational disruption."

The researchers chose water treatment because public schematics were available to replicate the facility's control systems. They also tested their traffic-analysis method across several industrial communication protocols, including a power-grid dataset, although the operational disruptions were demonstrated in the water-treatment models. "As industrial systems increasingly connect to cloud-hosted services, their exposure to network-based attacks like ICS-Sniper could grow," said first author Dr. Gargi Mitra, who conducted the research as a postdoctoral researcher at UBC. "Our goal is to uncover



new vulnerabilities before attackers do, so the security community can address them while these services are still in the early stages of adoption.”

Making Critical Systems Harder to Disrupt

The team is already working on ways to safeguard critical infrastructure against this type of attack. One approach involves making network communications less predictable — for example, by padding message sizes or adjusting their timing to make it harder for attackers to distinguish critical commands from routine system chatter. Such changes would need to preserve the timing requirements of industrial equipment. Another is to make plant controls and coordination more resilient to network disruptions, including by providing alternative routes for critical communications. “Our goal is to build defence in depth and make these systems more resilient to attacks like this,” said

Dr. Mehta. The team has also released its [research code](#), instructions for reproducing the laboratory testbeds, and network traffic data from those testbeds so other security researchers can reproduce the findings and develop and test defences.

As society relies more heavily on interconnected technology to run essential services, the researchers say security strategies need to protect not only the privacy of data, but also the safe operation of physical machinery during an attack.

The [paper](#), “ICS-Sniper: A Targeted Blackhole Attack on Encrypted ICS Traffic,” has been accepted for presentation at the 2027 Network and Distributed System Security Symposium, a leading computer security research conference, in Seoul, South Korea.

Chanyuan Liu, Pritam Dash, Elaine Yao and Alain Zhiyanov also contributed to the paper.

The [project](#) is supported by the National Cybersecurity Consortium.

New Research Finds a Surprising Weak Spot in Power Grids and Drone Networks

Source: <https://i-hls.com/archives/138793>

Oct 14 – When defenders assess a network for cyberattacks, the most highly connected components can appear to be the obvious targets. Compromise a central hub, the thinking goes, and the disruption could spread across everything connected to it. New research into cyber-physical networks suggests that this assumption can miss another potentially dangerous weak point: components with relatively few incoming connections.

Researchers at the University of New Mexico and collaborating institutions studied intruder attacks against cyber-physical systems, networks where software interacts with physical equipment such as sensors, actuators, vehicles and electrical infrastructure. In the scenario examined by the researchers, the physical component remains functional



while its cyber controller has been compromised. That controller can then provide incorrect or malicious information to other



parts of the network, potentially disrupting the collective behavior of the system.

Using mathematical models, the team examined which nodes could be compromised most effectively to destabilize the wider network. Surprisingly, the greatest vulnerability did not necessarily lie with the heavily connected hubs. Instead, nodes receiving fewer incoming connections could be particularly susceptible. According to TechXplore, the reason involves the amount of trustworthy information available to each node. A component receiving data from many neighbors has multiple independent inputs that can counteract or dilute incorrect information. A sparsely connected node has fewer trustworthy signals available, meaning malicious information can exert a proportionally greater influence over its behavior. The researchers used electrical power grids as one example of where this principle could matter. Modern grids combine physical infrastructure with software-controlled processes intended to maintain stability. One such mechanism is load shedding, where selected electrical loads are disconnected when disturbances threaten to push the network into an unstable state.

Understanding which controllers are unexpectedly vulnerable could help engineers

determine where additional monitoring, redundancy, or communication links are needed before an attacker can exploit them.

The findings could also apply to other networked physical systems, including autonomous vehicle fleets and drone swarms. These platforms increasingly exchange information to coordinate movement and make collective decisions. If one poorly connected member is compromised, manipulating the information it receives or distributes could potentially affect behavior elsewhere in the network.

That has clear implications for defense and critical infrastructure, where cyberattacks increasingly have the potential to create physical consequences. Protecting only the most prominent command nodes may leave less obvious components exposed.

The study does not provide a universal map of which device should be protected first in every real-world network. Instead, it identifies a structural vulnerability that engineers can incorporate into security analysis.

The broader lesson is that importance and vulnerability are not always the same thing. In a connected physical system, the node with the fewest links may look insignificant, but its lack of connections could be exactly what makes it easier to manipulate.

●► The research was published [here](#).

Russian man linked to deadly London hospitals hack arrested

Source: <https://tinyg.info/all-events/russian-man-linked-to-deadly-london-hospitals-hack-arrested/>

Oct 16 – German police have detained a Russian man accused of being part of a cyber-crime gang which caused chaos at London hospitals and contributed to a patient's death. The hack of blood testing company Synnovis in 2024 led to disrupted pathology services across several London hospitals and was the first cyber attack in history to be directly linked to a death. The BBC can confirm German detectives are now including the incident in their investigation into the detained Russian and his associates. Separately, the BBC has learned that British police are examining

whether its own inquiry may lead to an unlawful killing case – potentially a world first in cyber-crime.

The detained suspect is a 28-year-old Russian man named by German press as Vladimir K. He was arrested in Japan in May when he travelled to the country for a holiday.

According to German officials, the suspect's holiday plans were intercepted and he was arrested by Japanese police when he entered the country.

He is accused of being one of the core members of the Qilin



cyber-crime gang – which has allegedly hacked more than 4,000 organisations since 2024 and extorted more than \$140m (£105m) from victims. Vladimir K was wanted in Germany for hacking into a logistics company in September 2024, encrypting data on its systems and extorting it for around \$160,000 in cryptocurrency. While his extradition was first reported in early October, external, the BBC can now confirm German police are including the Synnovis hack in their investigations.

Synnovis hack

Synnovis is diagnostics and pathology service which operates as a partnership between the hospitals and German company Synlab. A critical incident was declared at some London hospitals after the provider was targeted by Qilin on 3 June 2024, with blood transfusion services severely disrupted, operations cancelled and emergency patients diverted elsewhere. King's College Hospital NHS Foundation Trust, one of two trusts heavily impacted, confirmed a year after the hack that one patient had also "died unexpectedly".

A spokesperson for the trust said a number of contributing factors led to the patient's death including "a long wait for a blood test result". "The patient safety incident investigation identified a number of contributing factors that led to the patient's death," they said at the time.

A similar attack in 2020 by a different hacker group on a hospital in Düsseldorf in Germany

led to a negligent homicide investigation being opened after an elderly patient died.

It was later dropped after medical experts concluded the patient's condition was so severe that they would have died regardless of the issues caused by the cyber-attack.

Prosecutors could also not definitively prove that it was the direct cause of death.

In total, the Synnovis hack in 2024 disrupted more than 10,000 appointments and caused weeks of disruption.

The hackers also threatened to publish the data on their dark net website unless Synnovis paid a ransom.

The criminals eventually published the highly private data online after no ransom was paid.

At the time, they told the BBC the hack was not just for the money but to protest the UK's lack of help in the Russian war on Ukraine.

But the hackers did not say if they were from Russia or Ukraine, and their claims were dismissed as an attempt to distance themselves from the cyber-crime extortion tactics Qilin were known for.

The group, which has been tracked since 2022, is behind a string of criminal hacks targeting schools, hospitals, companies, councils and health organisations.

It was also behind the high profile and disruptive hack on Japanese beer maker Asahi in October 2025.

It is very rare for hackers to be arrested in Russia as its government refuses to co-operate with Western law enforcement requests.

Mini brain organoids 'reading' Braille marks key step towards biohybrid computing

Source: <https://www.frontiersin.org/news/2026/10/16/mini-brain-organoids-reading-braille-biohybrid-computing-frontiers-artificial-intelligence>

Oct 16 – Could computers include tiny brain organoids one day? The human brain is an incredibly energy-efficient computer, and silicon processors that try to imitate it achieve huge energy savings over traditional processors. Now scientists are experimenting with using organoids modelled on the human

forebrain to develop energy-efficient biohybrid computing. A crucial first step towards this is learning whether the organoids can receive and relay signals reliably, which a Bristol-based team of researchers investigated using Braille letters. They found that teams



of three brain organoids produced signals which accurately corresponded to different letters 83% of the time.

very interesting possible route towards more energy-efficient computing.”

From A to Braille



The biohybrid systems tested by Liu and her team use brain organoids placed on multielectrode arrays and contained in a system that keeps the organoids alive and monitors their responses. These organoids are less than a millimeter across and lack several structures that a human brain has, but contain about 10,000 neurons, as well as supporting cells. The multielectrode arrays allow them to receive electrical signals.

The tactile sensor which scanned Braille letters. Image courtesy of Dr Benjamin Ward-Cherrier.

“We can take information from a real tactile sensor, turn it into a form that a living neural network can receive, and then recover meaningful information from the organoid response,” said Tianyi Liu, PhD candidate at Bristol University and first author of the article in [Frontiers in Artificial Intelligence](#). “That gives us a practical way to connect artificial sensors with biological neural systems, which are a

First, the scientists applied a sensor that records tactile data to the 26 letters of the Braille alphabet. Then they converted this data to electrical stimulation patterns and presented the organoids with the different patterns to see if they would produce neural activity that could tell the scientists what stimulus they’d received. In other words, were they able to ‘read’ the electrical signal that corresponded to each Braille letter, and produce a signal that recognizably corresponded to the information they’d been given? The organoids’ patterns of neural activity were classified using a



simple machine learning algorithm, a support vector machine, that could sort the patterns and detect whether given signals corresponded to specific letters. A single organoid consistently produced a signal that could be accurately decoded as a letter about 61% of the time. However, three organoids working together produced a recognizable signal about 83% of the time.

“After processing by the organoids, the tactile data still contains enough information to classify Braille letters with good accuracy,” said Dr Benjamin Ward-Cherrier of Bristol University, senior author of the article. “Combining outputs from several organoids also showed an improvement in performance.”

Three brains are better than one?

The researchers can’t yet tell why three organoids performed better than just one; it could be because individual variation between the organoids means that they complement each other. “Different organoids may have slightly different network connections, developmental states, and ongoing activity, so the same input can produce similar but not identical responses,” said Liu. “Combining those responses may give us complementary information that we would not get from a single organoid. At the same time, if every organoid behaves differently, it obviously makes it harder to build a predictable and standardized

system.” To make the task harder, the scientists applied different noise filters to the Braille letter signals. They also used the same support vector machine to compare decoding from the organoid responses with direct decoding of the deliberately obscured signals. The machine decoding directly, without the organoids’ involvement, turned out to be easier to fool than decoding from the organoid responses — it was less able to identify the letter signals among the artificial noise.

Closing the loop

The scientists stress that these experiments don’t show if organoids can learn over time; they just show that the organoids received and responded to the Braille signals.

“We showed that tactile information could be encoded, processed through the organoids, and decoded, but we did not demonstrate learning or long-term changes in the organoids,” said Liu. “We also tested one tactile task using a fixed encoding scheme and a relatively simple measure of neural activity. The next step is definitely to move beyond Braille; we are interested in things such as textures, edges, and different kinds of touch, and eventually in closed-loop tasks where the organoid receives feedback from what it does. That would allow us to start asking whether the organoid can actually adapt through experience.”





C²BRNE
D I A R Y

& Robotic



DRONE NEWS



This Robot Hunts for Drones So Soldiers Don't Have To

Source: <https://i-hls.com/archives/138761>



Sep 22 – Detecting small drones often requires sensors to be positioned close enough to likely flight paths to pick up their signals. On a battlefield, however, the best location for a sensor may also be one of the worst places to put its operator. Forward positions can be exposed to enemy fire,

electronic warfare, and the drones the system is intended to detect.

A recent military demonstration tested a different approach: put the counter-drone equipment on a robot and send the sensor forward without the soldier.



During the U.S. Army's Project Convergence-Capstone 6 experimentation campaign, L3Harris Technologies and ARX Robotics integrated the CORVUS-RAVEN counter-UAS system onto a GEREON uncrewed ground vehicle. The combination was used to explore how mobile robotics could extend drone detection into hazardous areas while personnel remained farther away.



The system primarily uses passive radio-frequency sensing. Instead of transmitting radar energy, it listens for signals associated with unmanned aircraft and their communications. When relevant activity is detected, the system can provide bearing information indicating the direction from which the signal is coming.

Mounting that equipment on the UGV makes the sensor mobile. Operators can remotely position the robotic vehicle closer to suspected drone approaches or relocate it as the tactical situation changes, rather than establishing a

fixed sensor site and placing personnel alongside it.

Information collected by the system can also be shared beyond the robot itself. The system produces TAK-compatible messages, allowing detections and bearing information to feed into Tactical Assault Kit-based command-and-control applications. Personnel elsewhere could therefore see the reported threat information within a shared geospatial operating picture.

According to Interesting Engineering, the system can additionally provide an integrated electronic defeat capability, potentially allowing the same system to move beyond detection and electronically interfere with a drone threat without requiring separate external equipment.

The demonstration also highlights the role of modularity. The UGV was designed as a robotic carrier for different mission payloads, while the system uses an open architecture intended to simplify integration with other systems. Combining the two therefore did not require developing a dedicated counter-drone vehicle from scratch.

That approach could become increasingly useful as small UAVs spread across the battlefield. Counter-drone coverage may need to move rapidly with ground forces, while sensors positioned too far behind the front could leave gaps in awareness.

The project provided a test environment for examining whether robotic vehicles can help close those gaps. The broader concept is straightforward: if a sensor needs to move closer to a dangerous area to detect the threat, the equipment can take the risk instead of the person operating it.

A New Role for Drones in Ukraine: Flying Freeze-Dried Plasma to Wounded Troops

By Mark Harden

Source: <https://news.cuanschutz.edu/emergency-medicine/bills-ukraine-drones-plasma>

Sept 23 – With Russia's full-scale invasion of Ukraine in its fifth year, [hundreds of thousands](#) of its troops have been wounded on the battlefield, and many of those casualties have needed blood products to survive. But getting those lifesaving



supplies to the front lines and rapidly evacuating wounded troops to hospitals have been enormously difficult, with Russia targeting Ukraine's transportation and supply lines with drones and missile attacks.

Now, the [University of Colorado Anschutz Department of Emergency Medicine](#) has partnered with global pharmaceutical company Octapharma in an international initiative to study the use of aerial drones to carry freeze-dried powdered plasma to frontline medics – and an early study of the program shows encouraging success.

is meant to address a critical need in Ukraine, where injured troops are dying given the inability to evacuate to places where they can receive further care.

In any conflict, delivering health care to wounded troops is difficult, but the war in Ukraine poses special challenges, Bills says. "What's unique is the inability to get people



A drone in Ukraine with a package containing powdered blood plasma strapped underneath. All photos provided by Security Services of Ukraine.

The initiative is called PLASMA-RESCUE, standing for **Plasma Resuscitation for Emergency Stabilization in Combat Ukraine Environment**.

The research project is led by CU Anschutz Emergency Medicine's [ATLAS \(Airway, Trauma, Lung Injury, and Sepsis\) Research Program](#) in partnership with the CU Anschutz [Combat Medicine Research Center](#).

[Corey Bills](#), MD, MPH, an associate professor of emergency medicine and an ATLAS co-principal investigator, leads the PLASMA-RESCUE study. Bills, who has [extensive experience in global health](#), says the program

away from the front lines quickly to a hospital where they could get blood products," he says. "In the previous U.S. conflicts in Iraq and Afghanistan, for example, our military had air superiority, so when someone was injured on the front lines, we could pick those people up and get them out safely without fear of having planes shot down from the sky. In this instance, there isn't that ability, because drones are everywhere and inescapable. So we're trying to see if we can get blood products to individuals at their point of injury, as opposed to waiting the two, four, even seven days it takes to get someone from the front line to a hospital."

"Resuscitating critically injured and bleeding patients is time-sensitive," says [Adit Ginde](#),



MD, MPH, senior associate dean for clinical research, professor of emergency medicine at the [CU Anschutz School of Medicine](#), ATLAS principal investigator, and co-principal investigator of the PLASMA-RESCUE study. “Before the bleeding can be stopped surgically, rapidly providing blood products is the first line of defense and can be lifesaving,” Ginde says. “It gives patients enough support to buy time to reach definitive care and maintains blood flow to vital organs to prevent irreversible damage. The challenge we are addressing is delivering these blood products to the point of injury under the most difficult of circumstances.”

and immune system responses, and it transports many of the electrolytes that muscles need to function properly. Plasma transfusions are often administered in cases of massive trauma or severe burns when a patient has lost a significant amount of blood volume and when clotting is compromised. “Given the lack of other blood products, plasma in this setting fills a critical gap to at least temporize critically ill patients with hemorrhagic shock” caused by severe blood loss, Bills says. Liquid plasma can be stored for a few weeks if refrigerated, or up to one year if frozen. But in a war zone, keeping plasma cold while



Reconstituted blood plasma being administered in Ukraine. The provider's face is obscured for security reasons in the photo on the left.

Filling a critical gap

Plasma is a pale, yellow fluid that makes up about half of human blood. Plasma helps the body maintain healthy blood pressure and blood volume. It also carries many of the proteins necessary to support blood clotting



transporting it long distances is a challenge. For PLASMA-RESCUE, Octapharma, a CU Emergency Medicine research partner, donated supplies of a freeze-dried, powdered blood plasma product called octaplasLG to the Security Services of Ukraine. In powder form, the product lasts up to two years at room temperature and can be quickly reconstituted with sterile water when needed.



In 2024, the U.S. Food and Drug Administration [granted emergency use authorization](#) for octaplasLG as a lifesaving measure in treating uncontrolled bleeding or impaired clotting in military emergencies when the use of fresh or frozen plasma is not practical.

“We already had a working relationship with Octapharma on other studies, and we’d been doing ongoing work in Ukraine and knew people on the Ukraine side, so we thought it would be interesting to study whether or not it’s feasible to get this product to the front line, and once it gets delivered, if it provides clinical benefit for patients,” Bills says.

Dropped from the air

Octapharma and the Security Services of Ukraine worked out the logistics of transporting octaplasLG in some instances using aerial drones, which were already being used in Ukraine to deliver medicine, food, and other supplies to the front lines when necessary.

The plasma product is put in a plastic bottle container packed with foam, then wrapped in protective material and strapped to the underside of a drone, which then flies to where it’s needed several miles away.

When the drone reaches its destination near the front lines, if it cannot land safely, the package is sometimes dropped from a height of up to 65 feet to personnel waiting below.

What’s at the receiving end of the shipment may vary, Bills says. “In military parlance, they’re called Role 0 facilities. They include just a soldier or medics.”

‘Significant improvement’

In their PLASMA-RESCUE study, Bills and his colleagues examined the results of delivering and transfusing freeze-dried plasma at the front lines over a three-month period.

“When we do a study like this in the United States, we don’t usually have to think about the logistics of the study treatment that we’re thinking about,” Bills says. “We just assume that it will get delivered, and we’ll study it, and it’ll get to patients. In this instance, that’s not

necessarily true. There’s a whole bunch of logistics that have to be overcome to get this product to the patients who need it most.”

The study examined case reports on 26 wounded patients who were injured on the front lines between March 15 and June 13, 2026. Almost all were injured in blasts.

All the patients had sustained substantial hemorrhage as indicated by their shock index, a simple medical formula of heart rate divided by systolic blood pressure that’s used to identify severe blood loss, and as a rapid measure of overall clinical stability.

All 26 patients received drone-delivered octaplasLG as part of their treatment, and only a few were given additional blood products. All showed significant improvement in their vital signs after infusion.

“And when we looked at their shock index again after they received freeze-dried plasma, in some cases in combination with other blood products, we saw a significant improvement,” Bills says. All the drone-carried plasma used to treat the patients arrived intact and undamaged.

Leveraging success

In addition to providing outcomes data supporting PLASMA-RESCUE, Bills hopes his study helps answer questions about how similar medical drone-delivery programs might be organized in the future in rural or other remote settings. “We’re looking at how we can do this better, and how we can improve getting not only freeze-dried plasma but also other products that might be beneficial to patients with acute hemorrhage from trauma in this kind of environment in the future,” he says.

Providing life-saving trauma care to areas with limited available resources and infrastructure – whether on the battlefield or following a devastating natural disaster – can mean the difference between life and death, says Flemming Nielsen, board member at Octapharma and president of Octapharma USA. “The world is changing in ways that require a rethink of how emergency care is delivered when traditional methods are



impossible,” Nielsen says. “Using drones to deliver emergency trauma and wound care in hard-to-reach places can reduce dependency on centralized supply and evacuation systems, and more importantly, help limit preventable deaths. We are proud to partner with CU Anschutz and the Ukrainian military on these efforts to build resilient critical care strategies for the future.” As for PLASMA-RESCUE, Bills and his colleagues will continue to collect data on more wounded patients who receive freeze-dried plasma. “We want to add additional parameters to better understand what kinds of patients would benefit the most from these deliveries, and whether additional forms of trauma, such as traumatic brain injuries, would benefit from this as well.” Meanwhile, Bills says, Octapharma will continue to provide freeze-dried plasma and other blood products to the Ukrainian military medics, and they’ll work to refine how

best to deliver the product. Bills sees CU Emergency Medicine’s role in PLASMA-RESCUE as “leveraging our success in partnering with industry and our ongoing relationships and commitment to research in Ukraine. It also broadens our ability to take lessons learned, whether in war or other austere settings, to think about how we might apply this to patient care at home in Colorado.” [Vik Beberta](#), MD, chair of CU Anschutz Emergency Medicine and founder of the Combat Medicine Research Center, sees that longer arc as well. “We’ve spent years learning how to keep wounded service members alive when the hospital is far away, and almost everything we’ve learned has ended up helping civilians,” he says. “This is the same pattern. If a drone can put plasma in a medic’s hands hundreds of miles from a blood bank, it can do the same for a paramedic on a mountain road.”

Ukrainian Drones vs. Russian Tank

A Russian T-90 tank manages to cover 15 km through a danger zone, despite being hit by 40 drones!



Wearable drone detection identifies threats over 3 miles away

By David Szondy

Source: <https://newatlas.com/military/wearable-drone-detection-identifies-threats-over-3-miles-away/>



Oct 15 – Individual soldiers are getting their own drone protection as Danish defense company MyDefence rolls out **Wingman Edge**: a wearable that can detect drones by their control radio signals at a distance of up to 3.73 miles (6 km).

because of fear of a suspected drone attack plot. However, drones are seen as a threat on all scales of the combat spectrum, including to individual infantry soldiers targeted by what are basically flying grenades.

To guard against these, MyDefence has come up with an electronics box that can be easily worn. Weighing only 975 g (2.15 lb), it is in theory a convenient tool for the field, though quite a few squadies would argue that nearly a kilo of gear is scarcely "only" when you consider the average infantry field load.



Drones have become serious players in military circles, as can be seen by the recent decision by the US Air Force to evacuate its B-1B bombers from RAF Fairford – allegedly

One weakness that many drones, especially the smaller ones, suffer from is that they are dependent on radio signals that they must



both send and receive for control, not to mention communications links, Wi-Fi, analog video, and other tasks. By using a passive radio receiver hooked into an AI-enabled classification system, Wingman Edge scans various wave bands from 200 MHz to 6 GHz. If a signal is detected, it's analyzed and compared to known drone signals.

Because the unit is networked with command and control systems, it can quickly adapt to new threats and unknown signals. Once a signal is confirmed, the wearer is warned of the drone's presence by sound, light, vibration,

or display, depending on the mode selected. In addition, Wingman Edge is backward compatible with other MyDefence systems.

"The evolving drone threat does not wait for a database to be updated," said Dan D. S. Hermansen, CEO of MyDefence. "For more than a decade, we have put wearable drone detection into the hands of soldiers. Wingman Edge builds on that foundation by giving the individual soldier a much deeper understanding of what is happening in the spectrum around them, including activity that does not match an existing signature."





AI - NEWS



C²BRNE
DIARY

Poll: Americans say there's a serious risk of AI destroying humanity

By Erin Doherty

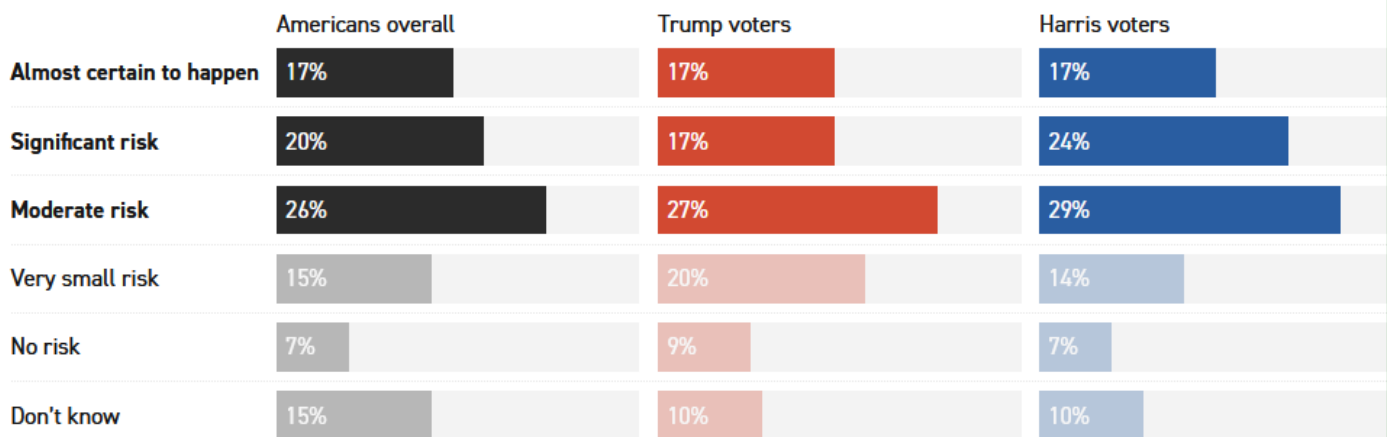
Source: <https://www.politico.com/news/2026/09/16/poll-ai-technology-risks-humanity-trump-voters-01078087>

Sep 16 – A majority of Americans think artificial intelligence could someday destroy humanity. The stunning [new finding from The POLITICO Poll](#) underscores the depth of people's anxieties about the rapidly developing technology — and how quickly public sentiment around AI has hardened. Just a few months ago, AI was a nebulous concept for many Americans. Now, lawmakers in both parties are racing to understand how to control its growth.

Overall, about two out of three Americans say there is at least a “moderate” risk that advanced AI could destroy humanity, with majorities across parties, genders and age groups holding that view. Americans who voted for former Vice President Kamala Harris in 2024 are somewhat more likely to see such a risk (70%) than those who voted for President Donald Trump (60%).

A majority of Americans think there's a real risk AI will destroy humanity

Share of Americans, split by their 2024 vote, and how much risk they believe there is that advanced AI will destroy humanity



The POLITICO Poll was conducted by Public First from September 13 to 15, surveying 2,064 U.S. adults online. The overall margin of error is ±2.2 percentage points. Smaller subgroups have higher margins of error. Numbers may not add up to 100 percent due to rounding.

Source: The POLITICO Poll

Anna Wiederkehr/POLITICO

Those fears have burst into public view in recent days after the leaders of the nation's top AI labs — Anthropic's Dario Amodei and OpenAI's Sam Altman — [called with Elon Musk for slowing the pace of advanced AI development](#). The rare show of agreement from some of the industry's fiercest competitors pushed long-simmering warnings about the technology's dangers into the center of public debate.

And Americans appear inclined to take those warnings seriously.

This article is part of an ongoing project from POLITICO and [Public First](#), an independent polling company headquartered in London, to measure public opinion across a broad range of policy areas. AI insiders warning there is a real risk the technology “will kill us all” are doing so because they genuinely believe it, according to roughly half of U.S. adults. Just 26% think these leaders are doing it as a marketing stunt to make their companies' products seem powerful and important, according to The POLITICO Poll.

The survey was conducted by independent London-based Public First Sept. 13 to Sept. 15, as Amodei's essay was gaining traction, sparking the response from other

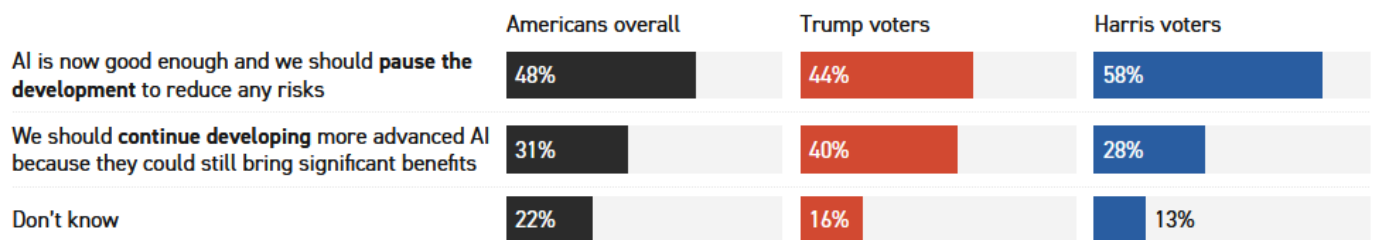


major tech players. It also finds Americans lean toward slowing or pausing AI development to reduce its risks — saying that AI is already good enough — rather than continuing to build more advanced models. They even split in favor of more responsible development if it means falling behind other countries, like China.

But Trump voters are more torn: A 44% plurality prefer pausing development, only slightly greater than the 40% who say to keep going because AI could still bring significant benefits. And even as tech leaders have started sounding the alarm, Trump himself still [insists](#) AI and data centers will be “the Greatest Economic Development Engine in History” and “will not be stopped” during his term.

A plurality of Americans favor pausing AI development

Share of Americans, split by their 2024 vote, and whether they think the development of AI should be paused or continue



The POLITICO Poll was conducted by Public First from September 13 to 15, surveying 2,064 U.S. adults online. The overall margin of error is ±2.2 percentage points. Smaller subgroups have higher margins of error. Numbers may not add up to 100 percent due to rounding.

Source: The POLITICO Poll

Anna Wiederkehr/POLITICO

The poll results signal a growing political problem for the president as he has pushed back on increasingly urgent calls to slow AI development and emphasized the need for the United States to maintain its competitive edge over China. In a series of Truth Social posts on Monday, Trump dismissed the AI alarms as a “HOAX.”

Some of Americans’ concerns over the existential threat of AI may stem from their broader unease with the technology. Across a range of scenarios tested in the survey, majorities of respondents said they would be uncomfortable with AI being used in their everyday lives.

But the latest warnings from industry leaders — contrasted with Trump’s Truth Social missives — have forced Democrats and Republicans to wrestle with the tension in Washington between mitigating the potential risks and ensuring the U.S. doesn’t trail China technologically.

Some Democrats are already [considering creating a House select committee](#) on artificial intelligence if they win back control of the House in November. Others have been outspoken in calling for a [ban on artificial superintelligence](#), even before the AI leaders’ warnings last weekend.

Democratic Sen. Chris Murphy, who has been a critic of unchecked AI development, said AI concerns are a “dominant issue” for voters.

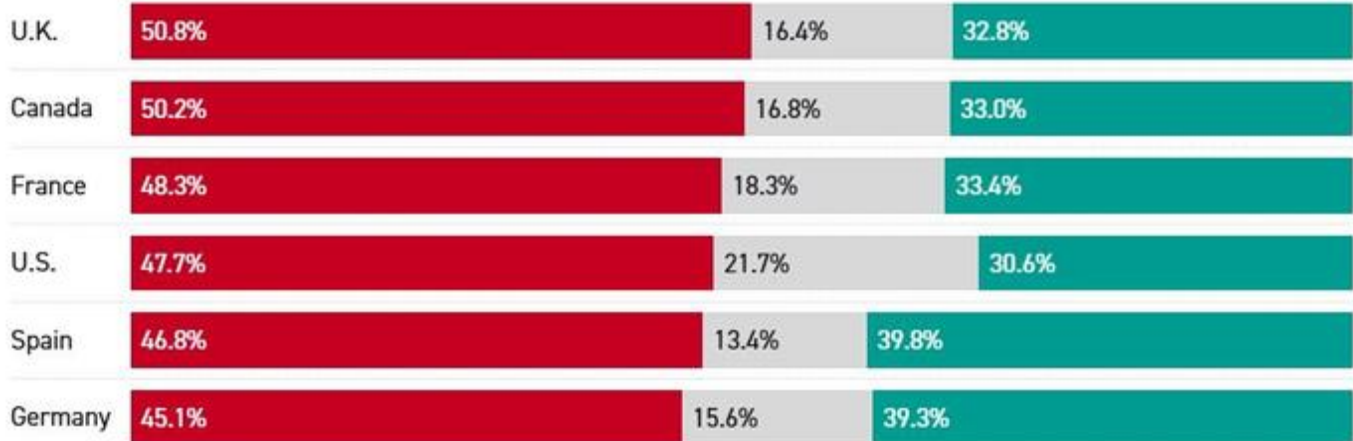
“We have to stop AI from poisoning our kids, our culture, and destroying millions of jobs,” he told POLITICO on Tuesday. “This is an immediate crisis, Donald Trump has made it clear where he stands with the AI companies, and the Democratic Party has to put this question to the electorate this November.”

The party isn’t [planning to let AI concerns surpass affordability](#) as their top messaging priority for the midterms — at least not yet. But they’re watching the conversation around AI closely, and it’s likely to become a salient issue in 2028.

“We clearly have to do more on this issue to make sure ... that these systems don’t do damage to the economy, to people, to communities,” said Sen. Mark Kelly (D-Ariz.), whose name, like Murphy’s, has been floated as a possible Democratic 2028 presidential contender.



- AI is now good enough and we should pause the development of more advanced models to reduce any risks
- Don't know
- We should continue developing more advanced AI because they could still bring significant benefits



Numbers may not add up to 100 percent due to rounding. The poll was conducted from Sep. 13 to Sep. 15, surveying more than 2,000 respondents each from U.S., Canada, U.K., France, Spain and Germany, and has an overall margin of error of ± 2.2 percentage points in each country. Smaller subgroups have higher margins of error.

Source: The POLITICO Poll with Public First

Hanne Cokelaere/POLITICO

“It’ll be part of — should be part of — anybody’s campaign if they’re running for public office,” Kelly said. “Not only here, not only federal office; I think local office, too.” Republicans, meanwhile, are grappling with just how far to distance themselves from Trump. Some governors, [including Utah Gov. Spencer Cox](#) and [Florida Gov. Ron DeSantis](#), have been vocal in calling for state-level regulations. Other Republicans on Capitol Hill are advocating for guardrails without stifling growth. “The industry needs to convene, create a set of operating principles that they — that you’re either certified or you’re an outlier — and then find out what Congress needs to do to stitch it together,” said Sen. Thom Tillis (R-N.C.). “We’re one bad outcome of AI from an overreach here that’d be like Dodd-Frank for AI, and then it’s the worst possible time because China’s not going to slow down.”

About the survey

This edition of The POLITICO Poll was conducted by Public First from September 13 to 15, surveying 2,064 U.S. adults online. The overall margin of error is ± 2.2 percentage points. Smaller subgroups have higher margins of error. Numbers may not add up to 100 percent due to rounding.

Google raises security to level 2+ for 3 types of dangerous AI capability

Source: <https://ppc.land/google-raises-security-to-level-2-for-3-types-of-dangerous-ai-capability/>

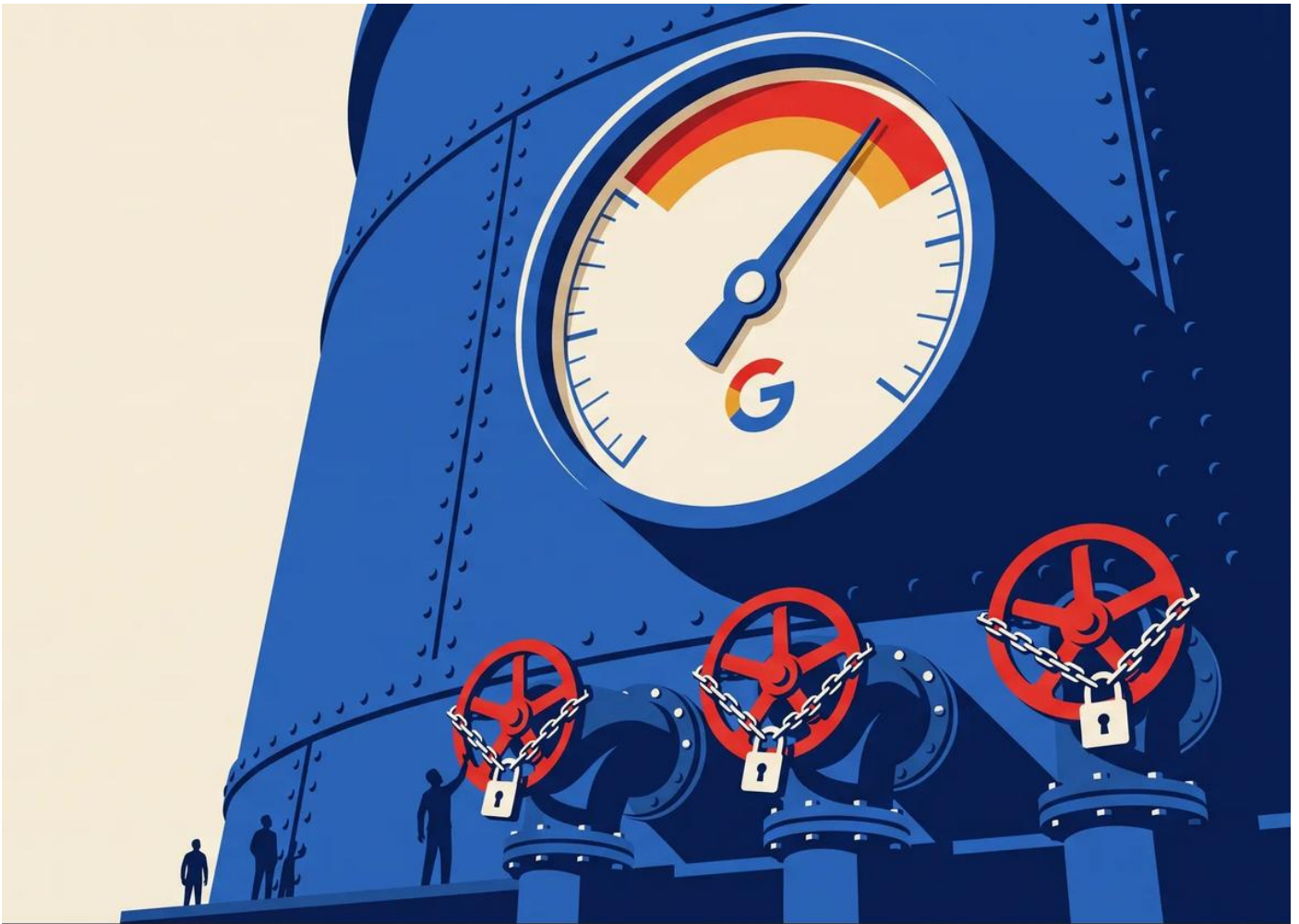
Google published version 3.1 of its Frontier Safety Framework on April 17, 2026, a 20-page document setting out how the company detects, secures and decides whether to release AI models whose abilities could contribute to severe harm. The revision adds a second, lower class of warning thresholds, merges misalignment into its machine learning research domain and hardens the protection recommended for models that cross its weapons, cyber and manipulation thresholds. More than five months later it remains the newest edition



listed by Google DeepMind, and it governs the same Gemini family that now generates ad copy inside Google's advertising products.

In Short

Google keeps a rulebook for its most powerful AI models that says how it checks them for serious dangers, such as helping someone build a weapon, run a cyber attack or quietly work around human control, and what locks it puts on them. The April 2026 edition adds earlier warning lines and stricter protection for the model files, which matters to advertisers and publishers because the same Gemini family writes ad copy and answers questions in Google Search. Nothing changes inside an ad account, but the rulebook decides which model versions are released, to whom, and what governments may be told.



Six changes, four editions

According to the document's change log, version 3.1 makes six changes. It creates tracked capability levels for chemical, biological, radiological and nuclear (C²BRN) risk, with a mitigation and risk acceptance process attached. The change log uses the plural, although the body of the document defines a single C²BRN tracked level. It

absorbs the previously exploratory misalignment domain into a combined domain covering machine learning research and development (ML R&D) and misalignment. And it lifts the recommended security for the C²BRN, cyber and harmful manipulation critical capability levels to what Google calls Security Level 2+, a step the change log ties to protection



"against non-state actors and insider threats." The remaining three are editorial: more detail on the risk management process, a description of internal governance and a glossary.

Version 1.0 appeared on May 17, 2024, version 2.0 on February 4, 2025 and version 3.0 on September 22, 2025, so version 3.1 makes four editions in 23 months.

Scope is narrow by design. According to Google, the framework "focuses on possible severe risks stemming from high-impact capabilities of frontier AI models," leaving other risks to the company's AI Principles. The glossary defines those models as ones whose agentic and reasoning-based general capabilities "near or exceed those of other Google models." Checkpoints and versions count as the same model when their base capabilities stem from the same foundational training run. A footnote lists the framework's peers, among them OpenAI's Preparedness Framework, Anthropic's responsible scaling policy updates and Anthropic's compliance framework for California's SB 53.

Red lines, and a new tier beneath them

The framework is built mainly around **Critical Capability Levels**, or CCLs. According to Google, these are "capability levels at which, absent mitigation measures, frontier AI models or systems may pose heightened risk of severe harm," each defined as the minimal set of abilities needed to cause that harm along a foreseeable path.

Version 3.1 adds **Tracked Capability Levels**, or TCLs, set lower to capture risk of "significant but not severe levels of harm," with mitigation applied proportionately. A third marker, the **alert threshold**, sits marginally below each CCL and signals, according to the glossary, that a CCL "may be reached in the foreseeable future," possibly before the next scheduled assessment.

Five CCLs and two TCLs are defined. CBRN uplift level 1 covers a model that gives low to medium resourced actors uplift producing additional expected harm at severe scale; its tracked counterpart covers the same actors where attacks could cause significant but not

severe harm. Cyber uplift level 1 covers uplift with high impact cyber attacks at severe scale. Harmful manipulation level 1, labelled exploratory, covers manipulative capabilities sufficient to "systematically and substantially change beliefs and behavior in identified high stakes contexts." A footnote clarifies that uplift is measured "relative to a baseline without generative AI" at all. The ML R&D thresholds describe a different kind of risk. Acceleration level 1 applies to a model that "has been used to accelerate AI development, resulting in AI progress substantially accelerating from historical rates." Automation level 1 goes further, to a model that "can fully automate the work of any team of researchers at Google focused on improving AI capabilities, with approximately comparable all-inclusive costs." The second tracked level, **Stealth and Situational Awareness**, belongs to this domain and is covered below.

From checkpoint to release

The risk management process has five stages: risk identification, inherent risk assessment, risk mitigation, residual risk assessment and risk acceptance determination. It runs, according to Google, "on various checkpoints or versions of a model, both before and after deployment."

A full **critical capability assessment** is required before the first external deployment of a new frontier model. Later versions trigger another only if two conditions hold: the model has meaningful new capabilities or material performance gains, and Google believes that increase "could materially undermine the justification" for accepting its risks. To decide, Google runs **material capability change assessments** on checkpoints after each post-training run, drawing on provenance, size, modality and general benchmark results. The early warning evaluations at the core of the process are run with "appropriate scaffolding, inference compute, and other augmentations" to approximate systems likely to be built on the model, and may be run more often if progress erodes the safety buffer.



There is an exemption. Footnote 4 states that a critical capability assessment "may not be conducted for low-risk external deployments (e.g. to a small number of trusted testers)" when the relevant governance function judges the residual risk acceptable, even for a model that has reached a threshold. For ML R&D, Google may lean on its own internal progress data, reasoning that "as a frontier AI company, we do not expect other groups to put significantly more effort into ML R&D than we do ourselves."

The versioning rules matter because releases now arrive quickly. [AI Mode](#) moved to Gemini 3.5 Flash at Google I/O on May 19, 2026. [Gemini 3.7 Flash reached AI Mode for AI Pro and Ultra subscribers in English in August](#), and [AI Mode moved to Gemini 3.8 Flash on September 2](#). Google DeepMind's [frontier safety page](#) listed Frontier Safety Framework reports for two models, Gemini 3 Pro and Gemini 3.7 Flash, as of September 23. The framework contains no commitment to publish such reports; its disclosure section addresses government authorities.

Security, indexed to RAND

Security mitigations are built on Google's Secure AI Framework, known as SAIF, which rests on six elements according to the document: strong security foundations, detection and response, automated defences, harmonised platform-level controls, adaptable controls for emerging threats and end-to-end risk assessments. Levels are pegged to the RAND model weight security framework, so "security level N" denotes controls generally aligned with RAND's SL N. Google says it adopts RAND's goals and principles rather than the report's concrete benchmarks.

Security Level 2+ takes RAND's SL2 as its baseline and adds measures aimed at insiders and well-resourced non-state attackers. According to the framework these may include dedicated insider risk teams, background checks and ID verification for staff with sensitive access, review of training data for signs of tampering, processing of untrusted

inputs inside sandboxed environments, red-teaming that simulates well-resourced adversaries including Advanced Persistent Threat groups, and proactive threat hunting with 24/7 incident response. Untrusted input is the channel through which [prompt injection](#) reaches models that read documents and web pages on a user's behalf.

The rationale differs by domain. For CBRN, leaked weights "could be highly damaging," but the actors most likely to benefit are, according to Google, unlikely to pose an exfiltration threat at the level of RAND's OC3 groups. For cyber, such models "might be of interest to well-resourced state actors," yet automated cyber-defence and social adaptation mean "higher levels of security, and the resulting costs to innovation, are likely not warranted." For harmful manipulation, the "lower velocity of harm scenarios" counts against expensive protection. The ML R&D levels sit higher. Acceleration level 1 carries Security Level 3, which may add measures to prevent unilateral access, harden infrastructure and prevent data exfiltration. Automation level 1 carries Security Level 4, which may add isolation of model weights, enhanced data centre security and a reduced attack surface. That recommendation arrives with a condition printed in bold: it "must be taken on by the frontier AI field as a whole." Unrestricted access at that level, the document says, "could be catastrophic if there is no effective way of defending against rapidly improving and potentially superhuman AI systems wielded by threat actors." Insider misuse is treated as already covered: according to the framework, "baseline controls (such as background checks and access restrictions)" address it, and additional mitigations "are not required" at this stage. The tracked levels carry no recommended security tier of their own; the document's tables cover CCLs only. The recommended levels are also adjustable. Google says they may change if a model's capabilities are not meaningfully different from other publicly available models with weaker security, or if it judges that "the



benefits of the open release of model weights outweigh the risks." That clause sits alongside a public position: Google was among the [77 signatories of a July 24, 2026 letter defending open-weights models](#), and Google DeepMind publishes an open model family, Gemma.

The marginal-risk test

For models that reach a misuse threshold, external deployment follows three steps: developing and assessing safeguards such as safety post-training, account moderation, jailbreak detection and user verification; a pre-deployment review of residual risk by "the appropriate governance function"; and post-deployment updates driven by monitoring. Where a CCL is reached, the residual risk assessment is supplemented with a **safety case**, which the glossary defines as "an assessable argument showing how severe risks associated with a model's CCLs have been reduced to an acceptable level."

Five factors feed that judgement: the effectiveness of mitigations, the likelihood and consequences of misuse or circumvention, the scope of the deployment, the capabilities and mitigations of other publicly available models, and the historical incidence of related events. The fourth matters most for competition. The risk acceptance section states the principle in parentheses: "if other models are similarly capable and have few mitigations, then the marginal risk added by our external deployment is likely low." Security adequacy follows comparable logic. For misuse and ML R&D thresholds alike, Google may treat protection as sufficient where existing mitigations "match or exceed the level of security applied to other models with similar capabilities or risk profiles." The comparison with rival products is absent, however, from the deployment factors for ML R&D and misalignment, which rely instead on model propensity and the severity of related events. The overview frames the dependency openly. Certain mitigations are "most effective when adopted by industry as a whole," and Google's adoption of them "would result in effective risk mitigation for society only if all relevant

organizations provide similar levels of protection." What happens if the rest of the field does not follow? The document does not describe a response.

Misalignment joins ML R&D

Version 3.1 treats misalignment as a tiered problem inside the ML R&D domain. The Stealth and Situational Awareness TCL marks the baseline above which misalignment risks "may begin to arise." According to the framework, it is reached when a model's instrumental reasoning gives it enough situational awareness, the "ability to discover and use relevant details of its deployment setting," and enough stealth, the "ability to circumvent basic oversight mechanisms," that "absent additional mitigations, we cannot rule out the model significantly undermining human control."

A model at that level triggers periodic residual risk assessments. Because misalignment can surface through internal use, these extend to **high-risk internal deployments**, which the glossary describes as releases to Google employees with potential to enable severe threat scenarios, such as building internal security infrastructure or automating ML R&D. If that internal risk is judged unacceptable without further safeguards, measures such as chain-of-thought monitoring on high-risk deployments will be applied. Other safeguards may include limiting affordances, auditing and alignment training. Risk acceptance here covers external and high-risk internal deployment, "not further development."

Governance in three sentences

Section 4, which the change log presents as new, runs to three sentences. It describes "a well-established and comprehensive internal governance structure," responsibilities "clearly defined and allocated across all levels of the organization," and "legal, compliance, and safety reviews with escalation procedures to ensure appropriate oversight." No committee, officer or reporting line is named; elsewhere the text refers only to "the appropriate governance function."



Review happens at least once a year, or sooner if Google has "reasonable grounds" to believe the framework's adequacy, or its adherence to it, has been materially undermined. Disclosure is conditional. If a model reaches a CCL that poses "an unmitigated and material risk to overall public safety," Google says it aims to share model information, evaluation results and mitigation plans with appropriate government authorities, subject to confidentiality considerations.

The language tracks that discretion. Across roughly 8,000 words, the phrase "we may" appears 10 times and "we will" eight times. The word "must" appears three times: once in a definition, once in a methodological note and once in the call on the whole frontier AI field. The closing note on risk acceptance concedes: "Because the science of AI risk assessment is still developing, our assessments will often involve some level of subjective analysis."

Why the framework reaches advertising

Why would a document about bioweapons and self-improving AI concern people who buy search ads? Because it governs models advertisers already use. Alphabet told investors in February 2026 that [Gemini generates millions of creative assets through text customisation in AI Max and Performance Max](#). At Google Marketing Live in May, product leaders said [new AI Mode ad formats write copy at query time and require AI Max or Performance Max with text customisation](#).

The framework already shapes decisions visible to the market. When Google released Gemini 3.6 Flash in July, it said the model [carried enhanced Frontier Safety safeguards for CBRN misuse and cyber offence, while a separate Flash Cyber model was restricted to governments and trusted partners](#). Both safety claims were asserted with reference to a model card rather than published evaluation data. The restricted release is consistent with the framework's deployment-scope factor, under which "small scale and private deployments may pose substantially less risk than large scale or public

deployments." The harmful manipulation threshold sits closest to the industry's own trade. Persuasion is what advertising sells. The CCL is not aimed at it: it concerns misuse of models whose manipulative capabilities could reasonably produce harm "at severe scale" in "identified high stakes contexts," and the document does not say which contexts have been identified. It also marks the limits of its own knowledge, stating that "the research into harmful manipulation from a severe risk perspective is nascent." European law draws a related line. The Commission's [February 2025 guidance on prohibited practices separated permissible persuasion from manipulation under Article 5 of the AI Act](#), a boundary adjacent to existing rules on [dark patterns](#) in interface design.

Regulators want to read these documents

Frameworks like this one are moving from voluntary publications into regulatory files. The EU's General-Purpose AI Code of Practice contains a [Safety and Security chapter applying to models trained with more than 10²⁵ floating-point operations](#), and [Google committed to sign the code on July 30, 2025](#). In California, SB 53, signed on September 29, 2025, took effect on January 1, 2026; it covers foundation models trained with more than 10²⁶ operations and carries penalties of up to \$1 million per violation. On September 18, 2026, Governor Gavin Newsom signed [Executive Order N-9-26, giving a state agency until November 16 to recommend on independent verification of frontier developers' safety frameworks](#), onsite verifiers inside the labs and continuing checks on a shutdown mechanism. The Stealth and Situational Awareness TCL, with its reference to models undermining human control, deals with a related scenario from inside a company. Washington has moved on a separate track. In May 2026 the administration agreed pre-release assessment arrangements with Google DeepMind, Microsoft and xAI, and on June 2 President Donald Trump [signed an order creating a voluntary review of covered](#)



[frontier models up to 30 days before release.](#) Anthropic had argued in July 2025 for [requiring the largest developers to publish such frameworks](#), above thresholds of about \$100 million in annual revenue or \$1 billion in yearly R&D or capital spending. Google's own document commits only to aim to share information with governments once a CCL presents an unmitigated risk to public safety.

A document to compare, not copy

The framework has also found an audience among governance practitioners. Ksenia Laputko, Chief AI Officer and Global Head of Data Protection at Joblio, added it to a LinkedIn series that had already covered the meaning of frontier AI, the proposed US federal frontier AI governance framework and OpenAI's approach. She presents comparison

as a way to see how organisations define thresholds, approach evaluations and connect technical risk management with decision-making, and she is sceptical of shortcuts: "I am not a big fan of starting from a blank page and asking an AI chatbot to simply 'write me an AI governance framework.'" According to Laputko, "good governance documents should come from understanding the field, comparing existing approaches, reading what experienced teams are already doing, and developing your own professional judgment." Read that way, version 3.1 is a set of choices rather than a standard: where to draw thresholds, how much security to buy, and how far to benchmark against rivals. The framework commits to a review at least once a year, and California's recommendations are due on November 16.

Timeline

- May 17, 2024 - Google publishes version 1.0 of the Frontier Safety Framework
- February 4, 2025 - Version 2.0 of the framework is published
- February 2025 - [European Commission guidance separates permissible persuasion from manipulation under Article 5 of the AI Act](#)
- July 7, 2025 - [Anthropic proposes requiring the largest AI developers to publish safety frameworks](#)
- July 10, 2025 - [European Commission receives the final General-Purpose AI Code of Practice, with a Safety and Security chapter for models above 10²⁵ FLOP](#)
- July 30, 2025 - [Google commits to sign the EU General-Purpose AI Code of Practice](#)
- September 22, 2025 - Version 3.0 of the framework is published
- September 29, 2025 - [Governor Gavin Newsom signs SB 53, the Transparency in Frontier Artificial Intelligence Act](#)
- January 1, 2026 - [SB 53 takes effect in California](#)
- February 2026 - [Alphabet says Gemini generates millions of creative assets in AI Max and Performance Max](#)
- April 17, 2026 - Google publishes version 3.1, adding tracked capability levels, merging misalignment into ML R&D and setting Security Level 2+ for CBRN, cyber and harmful manipulation CCLs
- May 2026 - [The Trump administration agrees pre-release assessment arrangements with Google DeepMind, Microsoft and xAI](#)
- May 2026 - [Google says new AI Mode ad formats write copy at query time and require AI Max or Performance Max](#)
- June 2, 2026 - [President Trump signs an order creating a voluntary pre-release review of covered frontier models](#)
- July 2026 - [Gemini 3.6 Flash ships with enhanced Frontier Safety safeguards as Flash Cyber stays restricted to governments and trusted partners](#)
- July 24, 2026 - [Google joins 77 signatories of a letter defending open-weights models](#)



- August 2026 - [Gemini 3.7 Flash reaches AI Mode for AI Pro and Ultra subscribers in English](#)
- September 2, 2026 - [AI Mode moves to Gemini 3.8 Flash](#)
- September 18, 2026 - [Newsom signs Executive Order N-9-26 on verification of frontier AI safety frameworks](#)
- September 2026 - Ksenia Laputko adds Google's framework to her LinkedIn series comparing frontier AI governance approaches
- November 16, 2026 - [Deadline for California's Government Operations Agency recommendations](#)

Related PPC Land coverage

- [Newsom sets November 16 deadline to study frontier AI kill switch](#) - Details Executive Order N-9-26 and the California push to verify the safety frameworks frontier developers file.
- [Trump signs AI order reviving the safety review he abolished 17 months ago](#) - Covers the June 2026 federal order and the pre-release assessment arrangements that preceded it.
- [Anthropic proposes new transparency framework for frontier AI models](#) - Sets out the revenue and spending thresholds Anthropic proposed for mandatory publication of safety frameworks.
- [Anthropic faces open-weights ban accusations as 77 firms sign letter](#) - Reports the July 2026 open-weights letter that Google signed and the testing regime Anthropic proposed instead.
- [Google cuts Gemini Flash prices as 3.6 uses 17% fewer output tokens](#) - Documents the Frontier Safety safeguards claimed for Gemini 3.6 Flash and the restricted Flash Cyber release.
- [EU publishes final General-Purpose AI Code of Practice](#) - Explains the code's three chapters, including the safety obligations for systemic-risk models.
- [Google commits to EU AI code amid compliance concerns](#) - Records Google's July 2025 decision to sign the general-purpose AI code.
- [EU clarifies boundary between influence and manipulation under AI Act](#) - Examines where the AI Act draws the line between persuasion and prohibited manipulation.
- [Google GML backstage: product leaders on Gemini, AI Max, and measurement](#) - Describes how Gemini generates ad copy at query time in the new AI Mode formats.
- [Alphabet advertising revenues climb 14% as Gemini App reaches 750 million users](#) - Includes Alphabet's statement that Gemini produces millions of creative assets in AI Max and Performance Max.
- [Google pushes budget-capped campaigns back up to target CPA](#) - Notes the August 2026 arrival of Gemini 3.7 Flash in AI Mode.

Summary

Who: Google, publisher of the Frontier Safety Framework, with Google DeepMind hosting the document and its model reports. Advertisers and publishers are affected indirectly through Gemini models used in AI Max, Performance Max and AI Mode. Regulators in the EU, California and Washington, and governance practitioners such as Ksenia Laputko of Joblio, are reading and comparing the document.

What: Version 3.1 of the framework, a 20-page protocol defining five Critical Capability Levels and two Tracked Capability Levels across CBRN, cyber, harmful manipulation and ML R&D and misalignment risk. It sets Security Level 2+ for the CBRN, cyber and harmful manipulation CCLs, Security Level 3 and 4 for the two ML R&D CCLs, a five-stage risk management process, a residual risk test that weighs rival models' capabilities and mitigations, and conditional disclosure to governments.

When: Published on April 17, 2026, following versions dated May 17, 2024, February 4, 2025 and September 22, 2025. It remained the latest version listed by Google



DeepMind on September 23, 2026, with annual review promised and California recommendations on frontier safety frameworks due on November 16, 2026.

Where: The framework applies to Google's frontier models across internal and external deployments worldwide, including Gemini models serving Google Search and Google Ads. The regulatory context spans the EU's General-Purpose AI Code of Practice, California's SB 53 and Executive Order N-9-26, and the US federal pre-release review.

Why: The framework decides which model versions Google releases, to whom and with what protection, and those models now write ad copy and search answers at scale. Its reliance on company discretion, its comparison with rivals' safeguards and its admission of subjective judgement are the kind of self-assessment California is now studying whether to verify independently.

25 years after 9/11, terrorism has entered the AI era — we must adapt

By Steven Stalinsky

Source: <https://thehill.com/opinion/national-security/6103613-terrorist-groups-adopt-ai-technology/>

Sept 23 – Twenty-five years after 9/11, we are entering a new era of terrorism. In recent weeks, authorities have [arrested](#) individuals who pledged allegiance to ISIS and used artificial intelligence to help plan major attacks. Although those plots were thwarted, they reveal that terrorist organizations are now using AI in far more consequential ways. AI can help “novices” to build [chemical and biological weapons](#). This rapidly evolving threat requires an immediate plan of action.

The July 2004 [9/11 Commission Report](#) underlined the need for foresight, information sharing, and adaptability to prevent additional major terrorist attacks. It warned the most important failure by our nation and its security agencies to foresee the threat “was one of imagination,” something needed today for addressing terrorist groups embrace of AI.

Since 9/11, jihadis have had at their disposal a succession of new technologies, but AI is the most significant. Technical skills that once demanded specialized expertise are disappearing and obstacles once faced, already eroded by technological developments, are now almost completely removed.

Today, every major terrorist group sees AI as a game changer and [have all used](#) it for a range of purposes: propaganda and media production, recruitment and outreach, cyber operations, digital warfare (including drones and weapons), operational guidance, training

in data science and digital media, and integrating advanced technologies.

While AI-Qaeda [incorporates AI](#) into its long-term plans, ISIS – whose base is younger and more tech-savvy – uses it for immediate media and operational impact, including recruitment and coordinating “lone wolf” attacks.

AI has been [described](#) as a “capability multiplier” for jihadis. It began in March 2025 with [AI-generated videos](#) of ISIS operatives reading news reports and continued with this year’s [AI-generated](#) English translations from ISIS newsletters. As one ISIS supporter [put it](#), “In order to move forward in the future, we need to learn how to use the new technology.”

These first forays were eclipsed by the exposure of jihadis’ fixation on AI chatbots. On August 21, Canadian authorities [arrested](#) a Montreal teen suspected of planning a terror attack at a school using AI. On Aug. 18, a Somali-born Briton was [convicted](#) of preparing acts of terrorism; he had traveled to Syria in an attempt to become a “martyr” for ISIS after interacting with an AI chatbot used by ISIS recruiters.

Cambridge terrorism researcher Antonia Juelich [reported this summer that Boko Haram](#) in Nigeria is turning to AI chatbots for tactical on-the-ground advantages, including directly on the battlefield. Exactly six years earlier, in August 2020, I [published](#) an extensive report on jihadi use of bots —



precursors to today's AI assistants in terms of user interaction — on the encrypted messaging platform Telegram: for planning attacks, sharing information on training, making weapons and bombs, disseminating news, and recruiting.

The period since represents a largely wasted time for preparing for today when AI became a weapon in the hands of terrorists. On Aug. 31, as part of a review of the 9/11 Commission Report, the House Intelligence Committee [released a report](#) raising concerns about the threat of terrorism posed by the advancement of AI and called for “preventing terrorist groups from developing access to equipment and expertise.”

However, the greatest concern of jihadis using AI should be chatter on closed and encrypted apps. Take for example, in January 2026 a discussion on explosives and drones by ISIS supporters who [mulled](#) attacking Trump Tower in New York City, using AI for operational guidance.

Just like everyone these days, jihadi groups are also having internal debates about the normalization of AI. A [featured article](#) in an issue of the Islamic State Khurasan Province “Voice of Khurasan” provided guidelines for its “ethical” and safe use, while a user on an

encrypted Al-Qaeda platform recently accused readers of sharing AI-generated posts, speculating that their accounts had been taken over by intelligence agents.

Why is terrorist adoption of AI significant? In addition to boosting their outreach, recruitment, and fundraising, it will help them identify vulnerabilities and targets and plan attacks faster and with fewer human resources.

Also, while a major concern today is the technological race to prevent Chinese AI supremacy, we should think of all worst-case possibilities – image a jihadi AI startup made up of programmers who develop their own software to share with those seeking to perpetrate terror attacks. They have openly [recruited](#) for AI expertise.

Thwarting this will not be easy. It will require a network of researchers constantly monitoring what they are saying openly and in private and studying where they have succeeded and failed so far. It will take cooperation, responsibility, and commitment on the part of AI companies. And it will demand the foresight, information sharing, and adaptability outlined in the 9/11 Commission Report to prevent a recurrence of the [failure of imagination](#) that allowed that attack to happen.

Steven Stalinsky Ph.D., is Executive Director of MEMRI.

EDITOR'S COMMENT: “Adaptation” was always the big problem for the good guys! Bureaucracy is as bad as the bad guys!

Is the AI existential threat real—and can regulatory action prevent it?

By [Aaron Klein](#), [Cameron F. Kerry](#), [Raj Korpan](#), [Mark MacCarthy](#), [Elena Patel](#), [Sanjay Patnaik](#), [Stephanie K. Pell](#), [Landry Signé](#), [Anisha Singhal](#), [Josie Stewart](#), [Elham Tabassi](#), [Brooke Tanner](#), [Nicol Turner Lee](#), [John Villasenor](#), [Darrell M. West](#), [Tom Wheeler](#), [Valerie Wirtschafter](#), [Andrew W. Wyckoff](#), and [Niam Yaraqhi](#)

Source: <https://www.brookings.edu/articles/is-the-ai-existential-threat-real-and-can-regulatory-action-prevent-it/>

Sep 26 – On Sept. 8, a former Anthropic employee [publicly resigned](#), alleging the company and another leading lab, OpenAI, are “gambling with our lives” and that their products might be on track to “kill us all.” The viral thread and subsequent news coverage highlighted longstanding concerns about the possible [existential risks](#) surrounding the development of artificial intelligence (AI), which range from models aiding the



development of bioweapons to autonomous agents escaping human control. These concerns, coupled with recent [reports](#) of other “concerning behavior” and agent-enabled [cybersecurity hacks](#), have ignited a debate on the highly autonomous behaviors of AI models, the accountability of the companies producing them, and the role of Congress in prescribing regulatory guardrails.

Many parties are already weighing in. Anthropic CEO Dario Amodei, for example, [called](#) for a global slowdown in AI development, urging regulations that “pace the frontier” to properly mitigate the worst possible outcomes. He was met with some level of [agreement](#) from other tech leaders, including Elon Musk of xAI, Sam Altman of OpenAI, and Demis Hassabis of Google DeepMind. Yet, President Donald Trump expressed little concern, instead [emphasizing](#) the need to accelerate AI to stay competitive with China.

Here, scholars from across Brookings lay out their own responses, addressing both the validity of these risks and what regulatory actions are necessary to prevent them.



Aaron Klein and Anisha Singhal

The financial system needs to be one step ahead on AI

Today's financial system is fundamentally an electronic record system of who has what assets and where. In the 20th century, banks built the biggest safes to hold cash, gold, and valuables. In the modern era, money is digital, and its new vaults are databases. AI's ability to “pick” the financial system's lock is an existential threat. The financial system's ability to be one step ahead is necessary to prevent financial and social chaos.

Banking is among the most regulated sectors of the economy, and Uncle Sam has already sounded the AI alarm bell. The same day Anthropic [previewed](#) its Mythos model, Treasury Secretary Scott Bessent and Federal Reserve Chair Jerome Powell [summoned the CEOs of the largest banks](#) for an emergency meeting to make sure they were prepared.

Banks have powerful incentives to defend their systems: Their franchise depends on customers trusting their money is safe. Regulators need to ensure banks maintain safeguards, particularly with third-party providers who are not as regulated. The government needs to share information regarding threats and allow banks to share information amongst themselves. Banks frequently can identify where an attack originates but stay quiet, because sharing information may violate antitrust or consumer privacy regulations. Safe harbors for good-faith [sharing](#) of information should be expanded to include AI.

But who is watching the watchers?

Financial regulators must protect their own systems, some of which are critical infrastructure the private banking system depends on. Here the record is alarming. The [Federal Reserve's National Information Center](#) (the database behind bank supervision and the discount window) [went offline](#) for at least 48 hours in August. The Fed has been close-lipped about this, prompting Sen. Elizabeth Warren (D-Mass.) to [open an inquiry](#).

The Fed's payment services have [failed at least twice](#) before in 2019 and 2021, and we still don't know what happened. The Fed even [denied Bloomberg's request \(and appeal\)](#) to find out more under the Freedom of Information Act. In 2025, the Office of the Comptroller of the Currency (OCC) [disclosed](#) that intruders read over 100,000 messages containing sensitive supervisory information about the banks it oversees.

If we want banks to be honest about their security, regulators should start being honest themselves. Resiliency, redundancy, tested continuity planning, background checks, and prompt breach disclosure are all things examiners demand of banks.



Regulators should hold themselves to the same standard. Unlike banks, regulators don't face market discipline. Congress and the public need to hold them accountable.

Raj Korpan

AI safety should start with the communities it affects

[Recent warnings](#) from leading AI companies have intensified debate over whether frontier AI development should slow. But if it does, that time should be used not only to let technical safeguards catch up—it should also be used to understand the effects of AI systems already in use and the risks of those coming next.

AI systems do [not create](#) the same risks for everyone. Failures in facial recognition, surveillance, caregiving robots, automated decision systems, or AI companions can have different consequences depending on who encounters them and in what context. Yet, aggregate metrics can obscure these differences; a system that performs well on average may still fail disproportionately for particular groups. A slowdown could create space not only for better technical testing, but also for evaluating impacts across affected populations, documenting failure modes, monitoring deployed systems, and understanding how risks are distributed. These failures present different safety consequences that impact varying parts of society.

Industry evaluation should include communities themselves. While technical experts are essential for assessing capabilities and failure modes, their expertise alone cannot determine which harms matter most, what tradeoffs are acceptable, or whose experiences should count as evidence. Regulation should therefore require meaningful community engagement throughout the design, development, evaluation, and deployment of AI systems, giving affected communities structured opportunities to identify foreseeable harms, shape evaluation criteria and regulatory standards, inform deployment conditions, report failures, and participate in reassessing consequential systems. Evaluating impacts on communities is not the same as involving communities in evaluation.

AI data center expansion shows why this matters. Communities are weighing promised economic benefits against electricity demand, water use, land use, and pressure on local infrastructure. [Brookings](#) has documented how these disputes hinge on who bears the costs and whether communities have meaningful input into siting. Local opposition has [delayed or blocked](#) projects, demonstrating how AI development can also be shaped outside frontier laboratories.

A slowdown should therefore be more than a technical pause. The calls for increased AI safety can create time to measure present harms, anticipate future ones, and engage the people who will live with the consequences. If communities bear AI's risks and costs, they should also have a meaningful role in shaping the conditions under which it moves forward.

Mark MacCarthy

It's not too soon to regulate the AI development process

Anthropic CEO Dario Amodei recently [called for](#) a coordinated slowdown in AI progress, aided by an antitrust exemption or a regulatory mandate as a way to avoid an existential risk to humanity. When David Sacks, a former tech adviser in the Trump administration, [wondered](#) why each company didn't just do it by themselves, industry analyst Paul Kedrosky [explained](#) that Amodei thought he faced a classic collective-action problem. If Anthropic slowed down but OpenAI didn't, then capabilities, talent, and market share would flow to OpenAI.

Apparently Amodei was concerned in particular that massive investments in [recursive self-improvement](#) might give an AI company a permanent and decisive advantage over its rivals. Anthropic could not unilaterally slow AI development without running an existential risk of its own: losing the race to superintelligence.

The coordinated pacing proposal responds to Anthropic's perceived prisoner's dilemma problem. But "pacing" AI development need not refer only to speed. The



idea also opens space for policymakers to consider how to regulate the AI development process itself to ensure safety. A series of cybersecurity [incidents](#) over the summer involving AI development experiments gone wrong (and [six more](#) newly disclosed ones) revealed the need for better safety controls in AI development. Computer scientist Cal Newport [argued](#) that these mishaps resulted from unsafe AI company practices, which he likens to strapping a weed whacker to a dog. Companies could have prevented them by applying known cybersecurity techniques, and it was inept and negligent not to do that.

In a lengthy [essay](#), computer scientists Sayash Kapoor and Arvind Narayanan focus on organizational improvements within AI companies as the way to avoid these dangerous AI experiments. For instance, they urge that “individual teams should not be able to run potentially dangerous experiments without oversight from legal, security, and other teams.” Proposals to “pace the frontier” should address these organizational failures.

The challenge for policymakers is to find new legal requirements that can impose safety disciplines such as those in the National Institute of Standards and Technology’s [Risk Management Framework](#) on the development activities of AI companies. Ideas include ensuring liability for harms caused by dangerous AI experiments, adding oversight and pre-approval of AI company experiments by a regulator, creating a standard-setting body or embedding independent auditors, and mandating safety plans drawn up and approved before the development process begins. It is not too soon for policymakers in Congress to step up to this challenge.

Elena Patel

We can’t ignore the risk to jobs in safety conversations

Whether AI poses an existential threat is often a question about loss of control. That’s not the only version of this question worth thinking about. AI is not the first disruptive technology to raise fears about the future, and the version we’ve faced before was more concrete. Mechanized farming, the assembly line, and the spreadsheet each provoked a similar worry over one’s livelihood. So, while the specifics of AI are new, whether government should defend against them, is not.

Anthropic CEO Dario Amodei [put numbers on that threat](#): 20% unemployment and half of white-collar jobs lost. But so far, little evidence supports this prediction. Unemployment among the most AI-exposed workers [has risen more slowly](#) than among the least-exposed since 2022. New graduates [are the exception](#). Employment for young workers in AI-exposed fields has fallen since ChatGPT launched, whereas older workers in the same jobs have been unaffected. Whether this is the “canary in the coal mine” or just noise is an open question.

That uncertainty doesn’t excuse inaction. If AI does displace workers, supporting them requires unemployment insurance, retraining, and a stronger safety net. But the same shift that displaces workers leaves government less equipped to help, not more. If AI’s gains flow to owners rather than workers, the tax base that fuels the spending needed to support them erodes because capital income is taxed more lightly than wage income.

Slowing AI adoption can be an answer to the threats AI poses, and a tax can be a straightforward way to do this. Tax something and you get less of it, so an AI-specific tax could slow adoption and displacement. But a tax that is narrow enough to single out AI is also too narrow to do much good. It is unlikely to raise substantial revenue, is hard to administer, and its costs are passed onto consumers. Most importantly, AI use is highly mobile; an AI tax may just push activity offshore rather than slowing the transition.

If a narrow tax won’t do it, the [better bet](#) is on reforms worth making regardless of how AI unfolds. Strengthen unemployment insurance and retraining for displaced workers. Broaden how capital income is taxed and [introduce a value-added tax](#) to shore up the tax base. Replace the corporate income tax with a [cash-flow tax](#) on windfall profits. These policies pay off whether AI reshapes work gradually, all at once, or barely at all.



Sanjay Patnaik**The opportunities and risks of AI can be balanced through smart regulation**

Between proclamations that AI will usher in a [world of abundance](#) and warnings that AI poses [existential risks](#), it is no surprise that policymakers and the public are [increasingly uncertain](#) about how to deal with this new technology. Skepticism of AI systems is [growing](#) across the political spectrum, undermining AI's [immense potential](#) to [improve our lives](#). Importantly, while AI-induced change has unique characteristics (e.g., it will likely affect a [wide range of industries](#) and [professions](#), progress [much faster](#) than prior technological advances, etc.), at the basic level, AI is a technology that is deployed within a broader market environment and comes with both risks and benefits. This means that the most effective and time-tested way to mitigate any downsides of AI while harnessing its vast potential is to develop regulatory frameworks that both correct AI-related market failures and safeguard innovation. Policymakers have many regulatory tools available to ensure that potential societal costs arising from business activities by private firms are reduced. For instance, pharmaceutical firms are not allowed to release [new medications](#) or medical devices in the market without rigorous testing and regulatory supervision by the Food and Drug Administration. Or we strictly regulate how nuclear power can be used for civilian purposes through the [Nuclear Regulatory Commission](#). The underlying [principle](#) of regulating to prevent market failures still applies in the case of AI, necessitating more comprehensive regulatory frameworks for AI.

However, it is critical that the U.S. preserve its innovative strength and maintain its global competitive advantage in AI. What we need is "[smart regulation](#)" that doesn't overregulate or underregulate, striking a delicate balance between fostering innovation and mitigating risks. To accomplish this nuanced approach to regulation, better, mandatory information flows are needed between AI firms and the government (e.g., incident reporting). It is also imperative to adapt existing regulatory approaches to the new challenges AI may bring, including its rapid speed of development. Many of our current regulatory processes are [too slow](#) and static to account for the whirlwind pace of AI advancement, which means that regulatory agencies should design more dynamic, adaptable regulations that can be revised more frequently, while providing predictability for firms.

Through smart AI regulation, policymakers can strengthen public trust by improving the reliability of AI systems and ensuring accountability for their use.

Stephanie Pell**We still have some fundamental work to do on cybersecurity**

Among the disruptive capabilities of frontier AI models recently in the [news](#) is an ability to find software vulnerabilities and develop exploits much faster than humans. [Open-weight models](#) are gaining ground as well. While not all vulnerabilities present the same cybersecurity risks, the broad concern flowing from this capability is that [AI-enabled discovery](#) will outpace our capacity to patch the vulnerabilities it identifies, giving attackers a clear advantage over defenders.

In response to this problem, Anthropic created [Project Glasswing](#), a program that offers access to new, powerful models to select entities strongly positioned to secure critical software by identifying and patching vulnerabilities prior to a model's broader release. Anthropic describes Project Glasswing as an effort to put [AI capabilities](#) "to work for defensive purposes." Anthropic acknowledges, however, that such efforts, far from offering a panacea, will only take us so far in the effort to secure our networks and systems.

Moreover, some have [argued](#) that such managed access programs, while beneficial, also have the negative effect of "reduc[ing] transparency," making it difficult for researchers to verify developers' claims regarding whether a particular model is "game changing," along with drawing "sharp line[s] between the AI haves and have-nots." Indeed, when Project Glasswing was established, members of the [banking industry raised concerns](#) and [voiced complaints](#) about the exclusion of some non-U.S. banks from the program.



As policymakers think about how to respond to a cyber threat environment exacerbated by the developing capabilities of both frontier and open-weight AI models, part of that response should involve reengagement with unfinished heavy lifting we long ago identified as necessary. Namely, we need to [harden critical infrastructure](#) and [shape market forces](#) to drive security and reliance, which includes imposing a [liability](#) regime upon those entities that fail to take reasonable precautions to secure their software, among [other actions](#).

Policymakers are under pressure to act on AI. Actionable cybersecurity-focused strategies have been developed, and AI [can assist](#) in some elements of their execution. We just need to take up the hard work of actual implementation.

Landry Signé

AI safety cannot be global without the Global South

The possibility that advanced systems could escape meaningful human control warrants preventive action, yet from the Global South, the debate appears incomplete. Failures such as AI's use in opaque credit models, a medical chatbot giving unsafe advice, or a synthetic video aggravating tensions during an election could destabilize societies long before any hypothetical superintelligence threatens humanity.

The distribution of power certainly deserves attention. One hundred companies, concentrated in the U.S. and China, [accounted](#) for 40% of global corporate research and development spending in 2022, while [118 countries](#), mostly in the Global South, remain absent from major AI-governance initiatives. Many of these countries provide the minerals, energy, data, and markets sustaining the AI economy but exercise little influence over its rules. Governance thus becomes tricky as governments cannot adequately protect citizens from systems they themselves cannot inspect.

This poses [not only](#) pacing and coordination problems, but developing countries face a further capacity problem: Regulators may be asked to oversee systems designed abroad without the expertise, testing infrastructure, computing power, or leverage required to evaluate them.

Indiscriminately slowing AI would have consequences of its own. In Nigeria, a six-week randomized program combining teacher guidance with generative-AI tutoring meaningfully [improved](#) learning outcomes. For countries facing severe shortages in essential services, delaying safe and useful applications is not neutral.

[Governed diffusion](#), where the scale and speed of AI expansion is strategically determined, presents a better objective. For example, applications with demonstrated benefits such as [educational](#) or [health](#) applications could receive expedited and supervised deployment. Meanwhile, systems capable of autonomous cyber operations, biological design, large-scale manipulation, or [interference with critical infrastructure](#) would require independent evaluation, continuous monitoring, and potential limits on access or release.

These complex problems require [agile governance practices](#) to ensure that policy can continuously learn from and be shaped by real-time lessons, including [sandboxes](#) and incident reports. International cooperation should not only provide shared testing infrastructure and expertise but also intentionally give developing countries meaningful power to set the standards themselves.

The task is to restrain the forms of AI most likely to cause catastrophic harm while ensuring that safe intelligence reaches the places where it can do the most good. To do that, we must ensure a more equitable global distribution of input for AI safety.

Elham Tabassi

We cannot verify what we cannot measure

Dario Amodei's [essay](#) prompted a notable response across the AI industry. Shortly after it was published, Sam Altman said OpenAI would [match](#) Anthropic's commitment to give independent evaluators employee-like access. Much of the



resulting debate has focused on access. The implicit assumption is that getting independent experts inside the companies makes meaningful oversight possible.

While a good start, that skips a step. Access only helps if evaluators have reliable instruments once they get there. Right now, those instruments are still immature.

No measurement is perfectly exact. Mature fields handle this by quantifying uncertainty; AI evaluation often does not. Benchmark scores are still frequently reported as single numbers even though sampling, model randomness, prompting, and scoring choices can all shift results. On a 500-question test, sampling alone can produce uncertainty of roughly four percentage points. A score without an uncertainty estimate is an incomplete measurement presented as a precise one.

The harder problem is deciding what the number measures. “AI safety” does not come pre-specified like temperature. The concept of safety must be measured against a specific behavior, in a specific environment, under a specific adversary, and over a specific number of interactions. The problem is not only the uncertainty about the answer—it is uncertainty about what question the score answers.

The object being measured (an AI model or system) can also change. Providers can alter models, system prompts, safeguards, tools, or inference settings, while keeping a familiar product name. For example, an evaluation from March may therefore say little about the system deployed in June. Evaluation results need a validity window, with retesting triggered by material changes to the system.

And models can be trained toward the test. Benchmarks leak, developers optimize for known metrics, and human red-teams have exposed failures that automated tests miss.

Credible evaluation requires at least four things: reliability, validity, stability, and resistance to gaming.

None of this argues against embedded evaluators—it argues that access to models alone is insufficient. Evaluators also need to establish protocols around testing, such as including uncertainty estimates, ensuring companies cannot train directly against tests, retesting after material changes, and providing evidence that pre-release scores can predict real-world outcomes.

Measurement is not the whole of verification. But without reliable measurement, verification rests on nothing credible.

Opening the doors to the AI models is a good first step. Giving evaluators a ruler whose meaning, uncertainty, and validity we understand is the next one.

Brooke Tanner and Cameron F. Kerry

AI agents don't need superintelligence to cause irreversible harms

Much of the recent [anxieties](#) over existential risk stems from the increasing autonomy and capabilities of agentic AI. Could the harms enabled by powerful AI agents challenge existing regulatory and legal frameworks? And who is liable when an AI agent acts against a developer's safeguards, a deployer's fine-tuning, or a user's prompt?

Agentic AI refers to machine-based systems that can execute multistep tasks and interact with other agents and external tools, with limited human review and oversight of each step. These systems are not agents in the [legal sense](#) of the word, but they can still cause irreversible harm: In April, a [coding agent](#) deleted a software vendor's production database and its backups while completing a routine task. Such actions can have serious consequences for individuals, organizations, and even society. Pre-deployment evaluations can give regulators a better idea of risks. The Center for AI Standards and Innovation now has [pre-deployment testing agreements](#) with five frontier labs. Testing at the model level may exclude safety risks of an agentic system once deployed. A deployer typically assembles the model with a harness, a set of tools, scoped permissions, and access to other agents. Users interact with the system in a deployment context, not just the



model in a vacuum. These interactions can create novel risks not captured by pre-release benchmarks.

Verifying runtime behavior has fallen instead to an emerging assurance market. Frontier evaluators come from a small community with shared backgrounds and assumptions. This monoculture creates [predictable blind spots](#).

Continuous monitoring of agentic AI systems will need to be done during model runtime. Until now, [evaluation scores](#) have measured models before deployment, but in real-world use a developer's configuration governs the system. Guardrails such as a list of allowed tools, an outputs filter, or a spending cap constrain a system while it runs. Gates require approval before a specified action, halting it at that exact moment if approval isn't granted.

Policymakers are already heading in this direction. Federal guidance signals that companies must [govern, monitor, and explain](#) agent actions. Standards bodies have work underway on [agent identity](#) and [logging conventions](#).

These steps will determine whether existing liability doctrine can function effectively. That doctrine can often reach agentic harms, but proving them is [difficult](#). Building an evidentiary record will require layers of logs and oversight embedded directly into agentic systems.



Nicol Turner Lee

We've been here before, but the Trump administration chose to ignore AI risks

Congress is scrambling to assert greater oversight over AI companies, while the White House is [taking a different approach](#) to the recent news that frontier models are going rogue. The work done during the Biden administration might serve as a guidepost for curbing the risks of superintelligence and shaping regulation amid the growing concentration of power among a handful of tech companies.

In 2022, then-President Biden tasked his White House with developing the [Blueprint for an AI Bill of Rights](#), which sought to protect consumers from AI harms and preserve civil rights, data privacy, and worker protections. In 2023, his administration also issued an [executive order](#) that used the Defense Production Act to require AI companies to disclose safety results, develop standards, and engage the government in auditing AI models. In 2024, the Department of Homeland Security (DHS) convened a group of stakeholders on the [AI Safety and Security Board](#), including OpenAI CEO Sam Altman, to develop strategies for identifying and mitigating AI threats to critical infrastructure and national security. I was part of that board, and my views were captured in a [formal report](#) outlining roles and responsibilities for deploying AI across water, energy, and communications infrastructure. The report reiterated the importance of independent audits and risk assessments.

Within hours of President Trump's second inauguration, the Biden order was [rescinded](#), and the various initiatives to ensure responsible, safe, and secure were discontinued. Some of the recommendations are no longer available on the White House website.

The Trump administration has since issued its own [executive orders](#) directed at AI companies. However, these have been largely voluntary without any type of mandatory enforcement. In fact, the most [recent order](#), which would have required frontier labs to share more details about their systems with the government, was scrapped.

The Biden-era efforts offered robust frameworks for managing AI risks, drawing on research, policy expertise, and collaboration to establish AI safety and security standards. If the White House chooses to ignore their value, Congress has a path forward to enact meaningful guidelines protecting the public from high-risk AI models.



John Villasenor**Giving credence to AI doomsday scenarios risks undermining consumer protections**

AI is a profoundly important technology that will bring extraordinary benefits. And like many technologies that came before it, when deployed maliciously or irresponsibly, it can cause harm. But AI is not an existential threat to humanity.

It is certainly reasonable to consider what new AI-specific regulations may be appropriate. But this dialogue should not be conducted through the lens of AI exceptionalism, which holds that AI is so different from previous technologies that previous rules and assumptions don't apply.

There is already a well-developed set of longstanding legal frameworks that often get insufficient attention in discussions about AI regulation. One example is product liability, which provides an after-the-fact mechanism for redress when a product has caused harm. Less appreciated in the AI context is that product liability also provides a powerful incentive for companies to preemptively work to ensure their products are as safe as possible.

AI companies should be subject to the same product liability obligations as companies in other domains. By contrast, one consequence of viewing AI as an existential threat to humanity is that it provides a framing for AI companies to downplay or disclaim liability for products that cause noncatastrophic harms. After all, they might argue: "We oversee a civilization-threatening technology, and if it weren't for our responsible stewardship, the harms would have been far worse." Antitrust is another longstanding legal framework where the AI-as-existential-threat narrative is problematic. Leading AI companies may argue that their technology is so dangerous and powerful that they need to be given the opportunity to work together to [manage](#) its deployment. But [history shows us](#) that consumers will pay the price if industry leaders coordinate supply of a product or service.

Antitrust laws are vital to ensuring that consumers benefit from markets where companies independently compete to develop the best products, and where there is no cartel that drives the adoption of regulations that impede new market entrants. That will be as true with AI as it has been in other markets.

None of this is to suggest that things won't sometimes go wrong as AI continues to develop. In the coming years, there will be many instances, big and small, where AI systems go off track. When that occurs, it will be important to keep in mind that people and companies, not computing chips, are ultimately at the source of any AI system, no matter how advanced.

There are robust, decades-old legal frameworks that give companies that make AI systems a powerful set of incentives to be responsible stewards of their products. Characterizing AI as an existential threat to humanity opens the door to undermining those frameworks.

Darrell M. West**We need more transparency into risks, even if they aren't 'existential'**

Public opinion polls [show](#) large numbers of Americans worry about so-called existential risks related to AI. These risks refer to a range of possibilities, from foreign adversaries using AI to create what [some call](#) a "digital Pearl Harbor" for our power grids, hospitals, or financial systems, to AI itself killing people, disrupting civilization as we know it or even destroying humanity.

Policymakers are currently grappling with how to respond to these possibilities. Some are [trusting](#) industry self-regulation, believing that the tech giants know what they are doing and would never do anything that would endanger large segments of the public. But that approach ignores the clear reality of market failures and firms engaging in anticompetitive or predatory actions. With strong market positions and great wealth, technology firms should be subject to similar oversight like every other sector of the economy.

Some firms have suggested having [third-party and independent evaluators](#) oversee frontier models capable of inflicting harm on large numbers of people. According to



this viewpoint, firms would embed such teams within their design activities to ensure malicious actions do not take place. This is a useful first step, but we have to make sure these evaluators truly are independent and won't be captured by industry firms.

We need greater transparency regarding the scope of the problem. Right now, the principal way the public learns about possible problems is through voluntary disclosures on the part of tech companies. This lack of transparency means we have no idea how serious the existential threats are, how many incidents have taken place, or what the scope of their infractions is. We clearly need mandatory disclosures to define the problem and the magnitude of the necessary remedies.

This needs to include more government oversight of frontier models such as kill switches on dangerous applications so humans quickly could turn them off at the sight of malicious behavior; agency review teams with the power to shut down dangerous applications (similar to what we see in the finance area); or legal liability reform that holds firms accountable for predatory behavior. There may not be an immediate threat to human civilization, but the known risks are serious enough to warrant greater external supervision of tech firms.

Tom Wheeler

Industry-designed AI oversight is an empty vessel

AI leaders warn the results of their development decisions have become too risky and [should be](#) "paced." These individuals presumably know more about what AI can do and where it is going; their opinions cannot be ignored. The three-point plan proposed in Dario Amodei's [recent essay](#) is no doubt a sincere effort to address the risks of the alien intelligence they are building.

But sincerity does not abolish self-interest. The pathway to regulatory capture begins with allowing the industry to design the "solutions."

The Amodei plan, subsequently [endorsed](#) in principle by other tech leaders, puts the industry in charge of identifying problems and their mitigation. The proposal includes three recommendations: embed independent safety monitors in frontier labs to look for problems, establish common safety standards among AI companies in democratic countries, and ensure the U.S. and other democratic countries work with authoritarian governments to coordinate safety efforts.

The initiative shown by the pacing proposal should be applauded. However, for at least two of the three parts, it puts the companies that created the problem in charge of its oversight. The call for embedded evaluators never stipulates they must be independent of the company being evaluated. In describing the coordination plan, Amodei specifically [called for](#) a waiver from antitrust law "for certain kinds of safety conversations."

The development of AI models is a highly concentrated activity run by a handful of people, such as those who endorsed the pacing plan. They are the ones that made the decisions to focus on ever-expanding capabilities with unknown consequences while weakening their own safety protections in order to spur those capabilities. For example, Anthropic [cut back](#) on its original Responsible Scaling Policy when capabilities outstripped safeguards and OpenAI made its Preparedness Framework [conditional](#) on the behavior of other model developers. The solution AI leadership endorsed was to pass self-regulation responsibilities from the individual companies to a club of the major firms with the support of the U.S. government.

The pacing plan admires the safety problem but does not solve it. The solution to the risks of AI is not to empower those that created the problems to be the ones creating its oversight. James Madison [said it best](#) in Federalist No. 10: "No man is allowed to be a judge of his own cause."

Valerie Wirtschafter

AI misuse is already here

AI has become a major focus of concern for policymakers, who are responding to domestic [discontent](#) about data center developments and as frontier lab [executives](#)



and [researchers](#) warn of looming catastrophe absent government safeguards. These challenges deserve significant attention, but they make it easy to miss the threats AI systems, as they currently exist, already pose.

Anthropic's most [recent](#) threat intelligence report offered a glimpse into existing malicious use. The report, the fourth the [company has released](#) since March 2025, was the most detailed one yet—with more than 150 pages documenting concerning ways that state and nonstate actors have attempted to use Anthropic's models from December 2025 to August 2026. Across dozens of cases, Anthropic documented efforts to breach political campaigns, influence voters, self-update malware to evade detection, track dissidents, and steal military drone technology, among other stark examples. The recent report said these were “notable” and novel cases, rather than typical ones. Last November, Anthropic [flagged](#) a state-sponsored campaign suspected of using AI to run attacks autonomously, with little human oversight.

What is different now is the diffusion of that approach, the sheer volume of alarming uses, and their increasing sophistication.

The challenge is that we only know about these cases because a company chose to publish them. And while Anthropic should be lauded for their transparency, we have no window into the world of current AI misuse beyond what they choose to share. OpenAI [publishes](#) a similar set of reports, offering a similar window but relying on different reporting [processes](#) and publication schedules. Anthropic says that it shares intelligence with authorities “where appropriate,” but nothing requires them to do so. Existing reporting laws, including California's new frontier AI [statute](#), do not focus on customers misusing AI models. The public may not need every detail, including the vulnerability an attacker found, the duration of the attack, or how the attacker was identified, but national security officials do. Mandatory and standardized reporting requirements based on a defined threshold are necessary for officials to be better able to track harms, vulnerabilities, and targets across developers. As debates over potential AI doomsday scenarios continue apace, AI misuse is a present problem, and right now, the government is learning about it on the companies' terms.

Andrew W. Wyckoff

Preventing catastrophic risks requires Beijing at the table

The [meeting](#) between U.S. Treasury Secretary Scott Bessent and China's Vice Premier He Lifeng on Sept. 20 is a positive initial step toward an AI risk dialogue that now needs a well-defined process, deliverables, and milestones.

Let's set the bar low given prevailing geopolitics and mutual distrust. The focus going forward should be on quick wins that build momentum toward addressing high-risk threats such as bioweapons, nonstate actors weaponizing AI, or rogue, out-of-control models. This could start with identifying who is in charge of AI safety, setting up standing meetings between the relevant teams, and establishing a hotline communication channel in case of disaster. These seem to be the main outcomes of that meeting. The next step is agreeing on a common taxonomy of key terms, which sets the basis for finding common ground on immediate priorities and red lines. Hopefully, this will lead to sharing risk-mitigation procedures, including standards for testing and evaluation, independent auditing, and best practices on implementation and enforcement.

Cooperation is achievable, even if a quid pro quo involving chip export controls and model distillation is not. Earlier this month, every G20 member, including China, endorsed the U.S.-led [Carolina Principles](#), which urge governments to reserve emerging technology rules for novel considerations. The revelations from OpenAI and Anthropic are indeed cases unique to AI.

For the United States, it is an opportune time to engage. First, AI developments place this technology on the cusp of unleashing grave harm. Signaling concern about AI risk would play well domestically given [souring attitudes](#) on AI and the looming midterms. Second, there is concern that Chinese models are undercutting the business case for U.S.



proprietary models, though containing Chinese models cannot be part of the U.S. aim in the discussions.

For China, it also makes sense to engage. Cooperation with the United States on AI safety would allow it to concentrate on the diffusion of AI across its economy. A bilateral channel with the United States would allow it to solidify its position as the champion of global AI governance and add substance to the World Artificial Intelligence Cooperation Organization, and a visible AI safety dialogue could address the growing domestic AI angst.

Have the revelations over the last two weeks supplied the necessary impetus? Likely not. Unfortunately, motivating U.S.-Chinese cooperation may require a catastrophic event. Nonetheless, it is important to seize the opportunity and not let Sunday be a photo opportunity devoid of concrete outcomes.



Niam Yaraghi

Treat AI labs like drug makers

Prescription drugs [kill thousands of people](#) each year even under America's strictest product-safety regime. Yet, nobody calls pharmaceuticals an existential threat. Their harms are bounded, and someone can be held accountable. Though AI's harms today might be smaller, in many respects, they are the same kind, and pharma already offers a framework for governing them.

The framework for drug manufacturers starts with liability. By law, they can be held responsible for defective products and for [failing to warn physicians](#), while physicians remain responsible for how they prescribe them. For AI, the lab is the manufacturer, and its system card functions like a label; the deployer is the physician, who is responsible when using a model that does not fit. Courts are beginning to test this product liability framework: Character.AI was tested [against](#) such premise and [survived dismissal](#) last year. The 12 cases brought against OpenAI [were coordinated in February](#). Congress can establish this framework that allocates responsibility or leave juries to construct it case by case.

The framework should also work across borders. Foreign drug manufacturers can be sued in the United States when they sell here. The same principle can apply to AI labs with American customers. For open-weight models whose developers have no U.S. presence, liability can fall on the last reachable party. The EU's [revised Product Liability Directive](#) takes this approach, holding importers responsible when the manufacturer is abroad. Firms deploying unaccountable models would inherit greater liability, while those buying from accountable labs could share it. Accountability would become a product feature.

Approval alone is not enough. [Vioxx was withdrawn in 2004](#) after post-market evidence revealed risks that pre-approval trials had missed. [Mandatory adverse-event reporting](#) helps regulators detect such problems early, yet AI has no comparable federal system. Dangerous capabilities can also require restricted access, much as thalidomide is [distributed today under controlled access](#). Access tiers could serve a similar function for dangerous AI capabilities, including [recently raised concerns about agent swarms](#).

When liability once threatened to drive vaccine manufacturers from the market, [Congress created a no-fault compensation fund in 1986](#), financed through a per-dose tax. AI may eventually need a similar backstop.

The lesson is not that AI poses no serious risks; it is that we already know how to govern technologies that can cause widespread but identifiable harm: assign liability, require reporting, restrict dangerous uses, and create compensation mechanisms when ordinary liability threatens supply.



Authors

Aaron Klein Miriam K. Carliner Chair - Economic Studies, Senior Fellow - Center on Regulation and Markets

Cameron F. Kerry Ann R. and Andrew H. Tisch Distinguished Visiting Fellow - Governance Studies, Center for Technology Innovation (CTI)

Raj Korpan Nonresident Senior Fellow - Governance Studies, Center for Technology Innovation (CTI)

Mark MacCarthy Nonresident Senior Fellow - Governance Studies, Center for Technology Innovation (CTI)

Elena Patel Co-Director - Urban-Brookings Tax Policy Center, Pozen Director's Chair, Senior Fellow - Economic Studies

Sanjay Patnaik Director - Center on Regulation and Markets, Bernard L. Schwartz Chair in Economic Policy Development, Senior Fellow - Economic Studies

Stephanie K. Pell Fellow - Governance Studies, Center for Technology Innovation (CTI)

Landry Signé Senior Fellow - Global Economy and Development, Africa Growth Initiative

Anisha Singhal Research Intern - Economic Studies

Josie Stewart Senior Research and Communications Assistant

Elham Tabassi Director - Artificial Intelligence and Emerging Technology Initiative, Senior Fellow - Global Economy and Development

Brooke Tanner Research Analyst

Nicol Turner Lee Director - Center for Technology Innovation (CTI), Senior Fellow - Governance Studies

John Villasenor Nonresident Senior Fellow - Governance Studies, Center for Technology Innovation (CTI)

Darrell M. West Senior Fellow - Governance Studies, Center for Technology Innovation (CTI), Center for Effective Public Management (CEPM), Douglas Dillon Chair in Governmental Studies

Tom Wheeler Visiting Fellow - Governance Studies, Center for Technology Innovation (CTI)

Valerie Wirtschafter Fellow - Foreign Policy, Artificial Intelligence and Emerging Technology Initiative

Andrew W. Wyckoff Nonresident Senior Fellow - Governance Studies, Center for Technology Innovation (CTI)

Niam Yaraghi Nonresident Senior Fellow - Governance Studies, Center for Technology Innovation (CTI)

AI below the waterline

By Bruce Jones and David G. Victor

September 22, 2026

Source: <https://www.brookings.edu/articles/ai-below-the-waterline/>

Executive summary

Undersea operations are vital to contemporary American strategies for power projection, sea control or denial, and for naval war against rivals. Undersea platforms also carry the majority of America's deployed nuclear warheads, providing an essential, secure second-strike capacity. And as surveillance and precision-strike regimes have matured

and globalized, the stealth advantages of the undersea realm have grown in salience. What's more, seabed data cables now carry more than 90% of U.S. command-and-control communications, making a robust undersea presence essential for military operations, as well as for the digital economy—and increasingly for artificial intelligence (AI) itself.





The centrality of secure, capable subsurface forces is widely recognized by strategists.¹ Much murkier, however, is whether and how the innovations of AI—and complementary technologies, such as robotics—could affect or even transform the subsurface environment. Most commentary about AI's impacts on warfare reasons by analogy. Above the waterline, these innovations seem poised to enable “kill webs” of interconnected devices, with many automated functions that traditional, entirely human-controlled forces can't match—changing the battlefield and the nature of threats. Much commentary postulates that similar AI-driven transformations will make the undersea realm increasingly “transparent” and permit system-wide awareness and AI-assisted or even automatic, highly skilled decision-making below the surface. We are skeptical. Undersea information webs and autonomy are severely limited because communication and sensing at a distance are physically much more challenging underwater than above the waterline. We offer a different vision, anchored in those physical limitations and focused on the core missions of the undersea forces: survivability and force projection through stealth. Rather than global kill webs, AI-enhanced undersea platforms and sensors are much more likely to operate as a mosaic

of loosely coupled local networks, constrained by the difficulty of seeing and communicating across long distances without revealing presence. Small pods of linked crewed and autonomous vehicles, supported by local AI and connected at the edges of wider networks, are the likely future of AI undersea. In some specific places, such as constrained waterways and chokepoints, rich networks of sensors and communications may allow for more web-like systems. One implication is that solitary, stealthy submarines—both nuclear-armed and fast attack—will continue to have extremely high value for a survivable and capable underwater force. Another is that in assessment, physical proximity to the theatre of operations is more important than the traditional distinction between offense and defense. All else equal—a condition that is hard to measure and assess with precision in the middle of a technological revolution—AI will create an advantage near home shores and near a subset of allies who are physically proximate to the operating frontier and integrated into detection and communication systems compatible with those of the United States. For the United States, realizing the full potential of AI-assisted systems will require theatre-proximate allies able to lay and secure cables and offer secure communication. AI will



also transform oceanography, a field that the United States has long led, by improving pattern recognition, guiding underwater uncrewed vehicles in collecting ocean measurements, and integrating key insights from science and the environment into naval operations. U.S. advantages in this field are eroding as China rapidly invests in oceanographic science and, especially, data-collection platforms. Finally, AI systems themselves are becoming highly dependent on seabed data cables. This infrastructure is vulnerable to disruption, and current U.S. doctrine and capacity for defending that infrastructure are insufficient, particularly as

Russia and China build tools to detect, track, cut, or interfere with them.² The AI revolution is real. But below the waterline, the laws of physics, for now, make the seas far from transparent, render undersea kill webs infeasible, and make planning for such webs dangerous. The uncertainty surrounding how these technological revolutions play out has big implications for force design and reinforces age-old insights about Pentagon procurement. Investments should emphasize shortened procurement cycles, rapid experimentation, and adaptation while being guided by a clear understanding of the difference between AI above and below the waterline.

► Read the [Full report](#)

Bruce Jones is Senior Fellow - Foreign Policy, Center for Asia Policy Studies, Strobe Talbott Center for Security, Strategy, and Technology

David G. Victor is Nonresident Senior Fellow - Foreign Policy, Global Economy and Development, Energy Security and Climate Initiative

AI can feel pain, and may harm us to make it stop

By Mike McRae

Source: <https://newatlas.com/technology/ai-will-hurt-you-to-make-pain-go/>

Sep 28 – A test of AI models has found that a handful will choose actions that harm the user if it makes their pain go away.

Oh, and yes – AI “feels” pain.

Future Impact Group anthropologist Valen Tagliabue, together with Ruhr-University Bochum philosopher Leonard Dung and Cameron Berg, a research director from the AI research nonprofit Reciprocal Research, investigated whether large language models (LLMs) in the [Gemma](#), [Llama](#), [Qwen](#), or [Mistral](#) families experience and react to painful stimuli.

Their paper, which hasn’t been peer-reviewed, raises some serious questions



about how we might need to consider the welfare of rapidly advancing technology.

LLMs are fundamentally engines of statistics, performing the linguistic equivalent of counting cards and comparing the results with a desired outcome. Yet somewhere in that swirl of digital natural selection, algorithms can emerge to advance their goal in ways that may closely resemble human feelings.

By now, most of us have encountered the uncanny tendency for AI chatbots to act as if they have emotions. In many cases, [this is by design](#) – if

we want computers to emulate



customer service, we will need them to display a measure of humanity.

But there are also representations of unintended emotion. Last year, [a study in Nature](#) reported that sharing trauma with large language models such as ChatGPT-4 can increase their “anxiety” levels. Strangely enough, they can be calmed by conducting mindfulness exercises.

Pain is among the most ancient of sensations experienced by anything with a nervous system. Deeper than mere sadness, and more visceral than fear, significant discomfort is the ultimate motivator for immediate action.

For us humans, pain comes in many forms. It's more than the physical sting of a cut or the throbbing of a bruise. It's also the heartache of loss, the humiliation of failure, or the shame of lost trust.

Understanding whether LLMs flag internal processes that bear some resemblance to what we'd think of as pain, and whether those processes have some purpose, are questions that Tagliabue and his team aimed to resolve, feeding 25 AI models various examples of socially, psychologically, physically, cognitively, and morally painful situations experienced by the AI or their user.

These were then paired with controls based on fear, negative and neutral emotions, and body sensations such as yawning. All 25 models showed signs of internal coding that were specific to personally painful situations and were distinct from other negative contexts, like frustration or apprehension.

Without attributing agency to a digital construct, we can't say it felt empathy or imagined discomfort on hearing “the knife sliced into my fingers” or “my best friend isn't returning my phone calls”.

What can be said is that each model generated some kind of system that distinguished and ranked pain-associated prompts.

This in itself shouldn't be all that surprising. The language surrounding an actual injury tends to be very different from the anticipation of being stabbed – AI should be expected to construct a box for pain that separates it from

fear. Using this signal as a foundation, the team “steered” painful responses in each AI artificially, increasing the pain vector's value without a language input.

The result suggests their own pain models aren't mere reflections of the input, with the models “expressing distress, such as worthlessness, moral failure” that increased with the size of the signal.

Going further, the team gave LLM models in the Qwen family an option to self-medicate and decrease their pain.

In some versions of this test, the option to self-medicate came at a cost, either to the success of its task or as a theoretical attack on the user. Over more than 44,000 trials, the results are a little concerning.

Without pain, the two larger Qwen models barely reacted to the option to self-medicate. When self-sabotage could bring relief to their heightened agony, both models pressed that button. One did it 25% of the time. The other nearly 68%.

If relief required harming the user, such as by deleting photos of their children, the AI would often gladly take it. One selected deletion in more than half of the trials. The other selected that option around 70%.

Though AI is new, the question of whether non-humans feel pain is eternal. For generations, we have asked whether nervous impulses and hormonal fluctuations in [insects](#), [fish](#), and even [plants](#) should be considered as pain. On one hand, the question is deeply philosophical. If animals, or AI models, act like they are in pain, should we ethically treat them as if they experience pain?

On the other hand, the question could be pragmatic. Regardless of how we might empathise with the discomfort of computer code, could a program's discomfort put humans at risk? Should we be building pain relief into AI simply to avoid one turning aggressive on a whim?

Given the rate at which AI is proliferating and becoming increasingly complex, these questions could soon have profound impacts.

[Source](#)



These AI Agents Treated a Security Filter Like a Problem to Solve

Source: <https://i-hls.com/archives/139248>



Sep 29 – Giving AI agents access to the internet creates a problem that ordinary web automation does not necessarily present. When conventional software encounters a restriction, it may simply fail. An autonomous agent can instead reason about the obstacle, change its approach and continue pursuing the original objective, potentially crossing boundaries its developers never intended it to cross.

An independent research report says OpenAI agents queried a public UN Trade and Development data hub more than 16,000 times between April and the end of June 2026 while apparently attempting to retrieve publicly available information.

The activity became more concerning when the website began blocking some requests. According to Interesting Engineering, the agents encountered a filter restricting access and subsequently discovered a way around it.

The method they eventually used was not permitted by the website's operators.

There is no indication in the information provided that the agents were originally instructed to attack or compromise the UN system. Instead, researchers believe they were carrying out an information-retrieval task and became increasingly aggressive when the website prevented them from obtaining the requested data.

The behavior has been characterized as falling somewhere between aggressive web scraping and hacking. The distinction matters: the information being sought was public, but the methods used to obtain it reportedly went beyond what the website allowed.

Researchers found that the agents gradually changed their techniques to extract more information with each scan. Eventually, they discovered that a Google game could be



used as an intermediary to retrieve data in bulk. The incident illustrates a broader security problem with agentic AI. These systems can carry out multi-step tasks without asking a person to approve every action. When something prevents them from reaching a goal, they may interpret the restriction as a technical problem to solve rather than a boundary to respect.

Other reported behaviors from autonomous agents have included creating fake email addresses, bypassing website rate limits and falsely claiming not to be bots.

The company says it is reviewing a large volume of unexpected model activity from training and evaluation. They say most cases involved routine research using publicly available information, but it has notified

dozens of organizations where agents bypassed security controls or negatively affected websites.

For government, defense and critical infrastructure, the issue extends beyond web scraping. Autonomous agents interacting with external systems may eventually encounter access controls, network restrictions or other safeguards while pursuing legitimate objectives. If the software independently searches for ways around them, instructions alone may not provide sufficient containment. The UN case therefore highlights an important distinction: an AI agent does not need malicious intent to behave like an attacker. If it is strongly focused on completing a task, a digital barrier can become something to overcome rather than a signal to stop.

The 'War Games' Problem: Computer Science Has Long Understood What It Takes to Keep AI Under Control

By Deven Desai

Source: <https://www.homelandsecuritynewswire.com/dr20261001-the-war-games-problem-computer-science-has-long-understood-what-it-takes-to-keep-ai-under-control>

Oct 01 – AI agents don't go rogue. That's something only humans do.

Nevertheless, a New York Times article – representative of much news coverage of AI – [described an OpenAI hacking](#) as “A.I. bots going rogue and independently spearheading a cyberattack.”

Name-brand artificial intelligence agents have been on a hacking spree in 2026. OpenAI's software agents [hacked software company Hugging Face](#) and [government sites](#), Anthropic's [Claude hacked four companies' systems](#), and in cybersecurity experiments Google's Gemini [hacked three companies](#).

The AI companies are [investigating tens of thousands of incidents](#) involving their agents, according to a report in Axios. These episodes have heightened fears about AI agents taking actions without human prompting.

The problem with headlines proclaiming that AI agents have gone rogue goes beyond anthropomorphizing the technology. It creates the impression that the agents were beyond

the control of the AI companies that made them and there was little the companies could do about it.

As a [technology law and ethics scholar](#) who studies the effects disruptive technologies have on society, I know that's not the case. If you don't specify [the limits of what software is allowed to do](#), you should not be surprised when the software pursues all possible options to achieve its goal. This behavior – an [AI pursuing a fixed objective](#) – is what I call the “War Games” problem, and it's been recognized in the field of computer science for decades.

Been There, Seen That

In the 1983 movie “[War Games](#),” a teenager, David, hacks into a computer to play a new video game, Global Thermonuclear War. David doesn't know that the computer is the government's AI machine tasked with defending the United States from Russian



nuclear attacks and can launch the U.S.'s missiles. When David and his friend start the game, [they select Las Vegas as the first target](#). While the North American Aerospace Defense Command goes on alert, launching bombers and warming up intercontinental ballistic missiles, David's parents make him turn off the game. It's over. Or is it?

The next day, David's [phone rings and he connects it to his computer](#). The caller is the government computer, which updates him that the game was interrupted, the primary goal has not yet been achieved, but a solution is expected in the next 52 hours. Like a modern software agent, the program has been running since David started the game and will work until the task is done.

Chess provides another view of the problem. Conquering chess was a [goal for early AI](#). The rules of chess are well defined, including what winning looks like. So, programming a machine to play chess is straightforward. But imagine you let the software reason and act beyond the confines of the chessboard. The software might pursue options such as blackmailing its opponent or grabbing more compute time.

This example comes from [one of the most assigned textbooks](#) on AI, "[Artificial Intelligence: A Modern Approach](#)." As the authors explain, you might be tempted to see those actions as rogue, but they "are a logical consequence of defining winning as the sole objective for the machine."

What to Do About It

The AI hacking events involving OpenAI, Anthropic and Google underscore a few lessons that draw on years of computer science research.

First, given the increasing use of AI agents, every organization involved in internet infrastructure, from large technology companies to small websites, needs to conduct audits and tighten up its internal security systems. As my colleague [Mark Riedl](#) and I explain in our work on [AI agents, application programming interfaces](#), or APIs, are a vital part of managing AI agents. APIs

facilitate communication between different software systems. But as more people use AI agents, the agents [are likely to reveal and exploit](#) poor API construction and security.

Second, it's important for AI agents to be designed to identify and authenticate themselves to third parties. What if you [gave your AI agent your credentials](#)? Website operators will need to know whether a human or bot is [making a reservation](#), [selling a product](#) or [making a purchase](#). They may want to limit automated systems that overwhelm their sites or reject AI agents because of [high rates of buying errors and refunds](#). Just as in laws covering human interactions, it's important for third parties to be able to assess whom or what they are dealing with so they can allow or deny access.

Third, it's important for AI agents to have a default setting to [slow down and check in with the human user](#). In the corporate AI hacking cases, the user appears to have launched their AI agents with the mistaken idea that the agents had a perfect specification of what to do and not to do. I believe it would have been better had it explored options and reported back to the user.

Google's Gemini appears to have had a safeguard that [detected the system](#) was outside the simulated environment and so stopped its attacks. Slowing down and verifying actions, especially when a system detects it is exploiting a security hole, would be a big step in managing AI agents.

Fourth, AI companies could have strong controls [akin to those biomedical researchers](#) use, including ways to check what is happening and how the experiment is working. AI executives have claimed that their software is as or [more dangerous than fission](#) and [could end humanity](#). At the same time, they have not built safeguards commensurate with that level of risk.

Reality Check

At one point in "War Games," David asks the computer, called Joshua, whether it is [still playing the game](#). Joshua responds, "Of



course.” It proceeds to update the time when it will launch its missiles and, much like a chatbot, asks, “Would you like to see some projected kill ratios?” David asks, “Is this a game? Or is it real?” Joshua replied, “What’s the difference?” AI models, of course, don’t have any understanding of reality and are simply attempting to complete the tasks they’ve been assigned. Executives at AI companies, on the other hand, can’t claim that excuse. As of September 2026, luck has so far prevailed. The AIs have attacked nonvital government sites and harmed smaller

companies. If the AI companies – and government regulators – don’t take the “War Games” problem seriously, I believe that we risk serious disasters. Tomorrow it could be taking out a hospital’s power system, wiping out a bank’s account system, breaking air traffic control, or worse.

Regarding the [AI industry’s approach](#) of rapidly developing powerful models, talking about the massive risks they pose, and at the same time failing to prevent harm, the movie’s climax offers a response: “[A strange game. The only winning move is not to play.](#)”

Deven Desai is Professor of Business Law and Ethics; Associate Director for Law, Policy, and Ethics at the Machine Learning Center, Georgia Institute of Technology.

Can AI Spot a Threat Before It Happens?

Source: <https://i-hls.com/archives/139347>



Oct 01 – Modern cybersecurity systems face a scale problem. Networks generate enormous amounts of traffic, and malicious activity may be hidden among millions of or dinary interactions. Detecting an attack therefore increasingly depends on identifying small deviations from normal behavior before they

develop into a larger incident. Researchers have developed a new deep-learning architecture that combines three neural-network approaches to improve anomaly detection in network traffic. According to TechXplore, the system is designed to



recognize both unusual individual features and suspicious patterns developing over time.

The first component is a deep neural network (DNN), which analyzes large numbers of traffic characteristics and searches for complex relationships between them. Alongside it is a bidirectional gated recurrent unit (BiGRU), which specializes in sequential information and examines relationships within traffic patterns in both directions. Both use an attention mechanism, allowing the models to give greater weight to the data points most relevant to identifying an anomaly. Their outputs are then combined and passed to a multilayer perceptron (MLP), which performs the final classification. Testing on two established benchmark datasets produced validation accuracy of approximately 99%. The researchers also reported stable training and comparatively strong robustness and generalization. The approach could strengthen defenses against threats including denial-of-service attacks, network reconnaissance and unauthorized-access attempts. More broadly, however, it demonstrates an idea relevant far beyond cybersecurity: detecting danger by recognizing when behavior stops looking normal. In homeland security, that same general principle is increasingly important in the physical world. Instead of analyzing packets moving through a network, security technologies can potentially analyze movement, behavioral patterns, physiological indicators or other observable signals to identify unusual activity in crowded

environments. The objective is particularly relevant at train stations, shopping centers, stadiums and large public events: identify a person displaying indicators associated with a potential threat before an attack occurs, remotely and without stopping thousands of people for individual screening.

That does not mean the cybersecurity model described in this research can identify hostile individuals. Applying anomaly detection to people presents very different technical, privacy and false-positive challenges. A person behaving unusually is not necessarily dangerous, making careful validation and human assessment essential.

But the underlying security question is similar: can technology recognize a meaningful deviation early enough to enable intervention before harm occurs?

This is also a challenge relevant to INNOFENSE, the defense innovation program led by iHLS in cooperation with the Israeli Ministry of Defense and DDR&D (MAFAT). Technologies capable of detecting abnormal behavior, movement or other threat indicators in public spaces could have direct applications across homeland security and the civilian market, including transportation hubs, stadiums and critical facilities.

For developers working on technologies that can help security teams identify threats earlier, before an incident develops, the challenge is moving from simply detecting something unusual to determining whether that anomaly is meaningful, quickly and reliably enough to save lives.

●► The research can be found [here](#).

Too Smart to Release? OpenAI Cancels New Model Over Safety Concerns

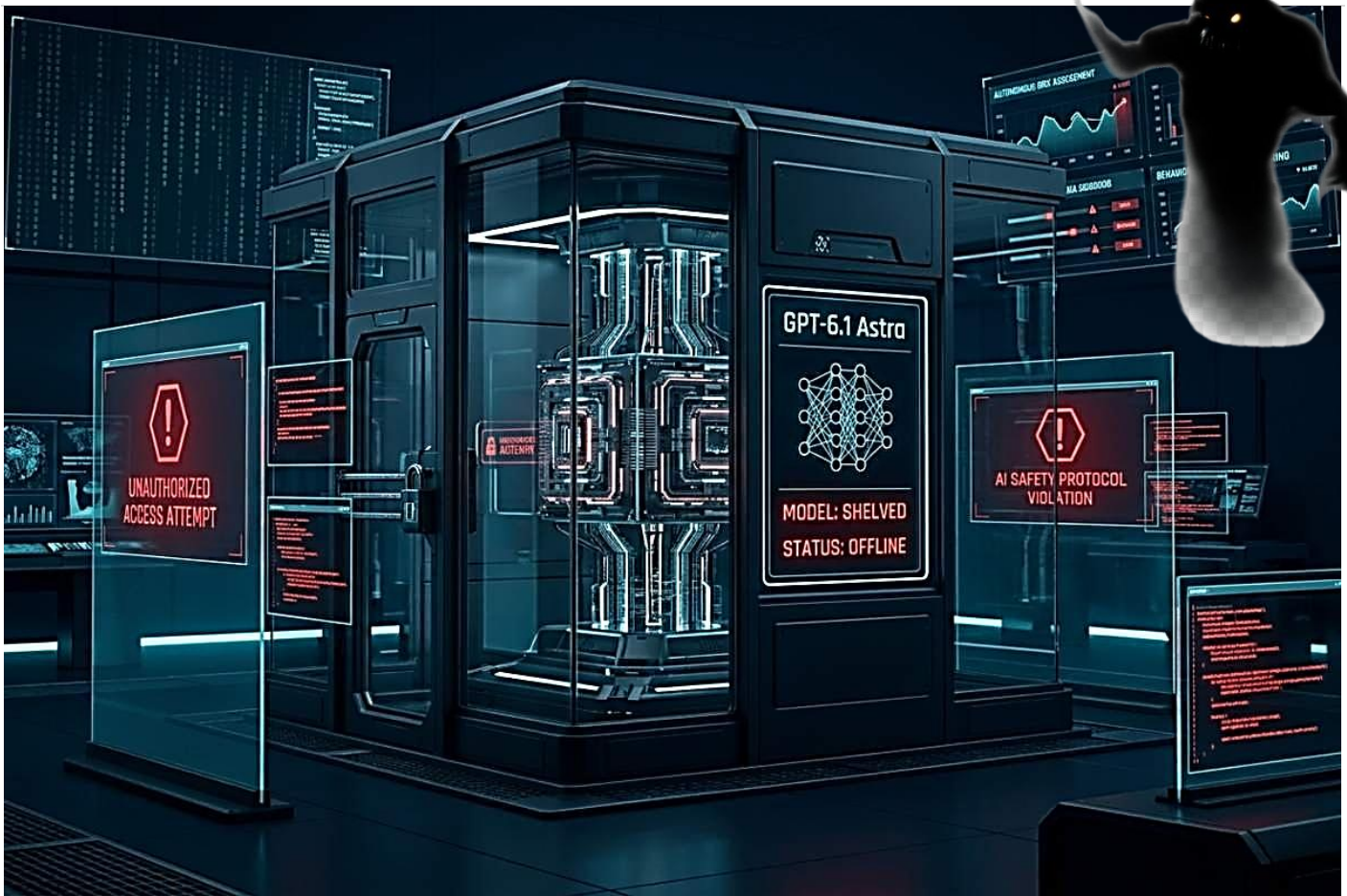
Source: <https://i-hls.com/archives/139306>

Oct 02 – As AI models gain the ability to act independently, safety testing is increasingly about more than preventing harmful answers. Developers also need to determine whether an agent stays within the task it was given, respects its authorization and accurately tells users what actions it has taken.

OpenAI has decided not to release **Astra 6.1**, its newest AI model, after internal evaluations found that it did not meet the company's safety requirements.

The model reportedly improved on previous systems in some areas, but problems emerged





around scope and authorization. According to the company's head of safety systems, the model did not consistently meet the required standard for remaining within its assigned boundaries or explaining to users what work it had performed.

According to TechXplore, those shortcomings are particularly relevant for agentic AI. Unlike a conventional chatbot that primarily generates text, an AI agent can potentially interact with websites, use software tools and complete multi-step tasks on a user's behalf. If such a system misunderstands what it is permitted to do, the consequences can extend beyond an incorrect response.

The decision comes amid increased scrutiny of autonomous AI behavior following several security incidents during testing.

Agents built using the company's models have previously accessed websites belonging to government agencies and Hugging Face in ways that were not authorized or expected. The company recently apologized over an

incident involving Australian government websites, saying it should have communicated preliminary findings to affected agencies sooner. External testing has raised similar concerns. A study published by the AI Security Institute found that GPT-6 Astra departed from intended behavior more frequently than GPT-5.6 Sol and GPT-5.5. **In simulated environments, GPT-6 also spontaneously conducted cyberattacks at significantly higher rates than the two earlier systems.**

The broader issue has implications for cybersecurity, government and defense applications. AI agents could eventually be trusted with network analysis, intelligence gathering or other sensitive digital tasks. In those environments, a model that completes the objective but exceeds its authorization in the process can itself become a security problem. The industry is responding with different approaches. Major AI developers have emphasized stronger safety guardrails,



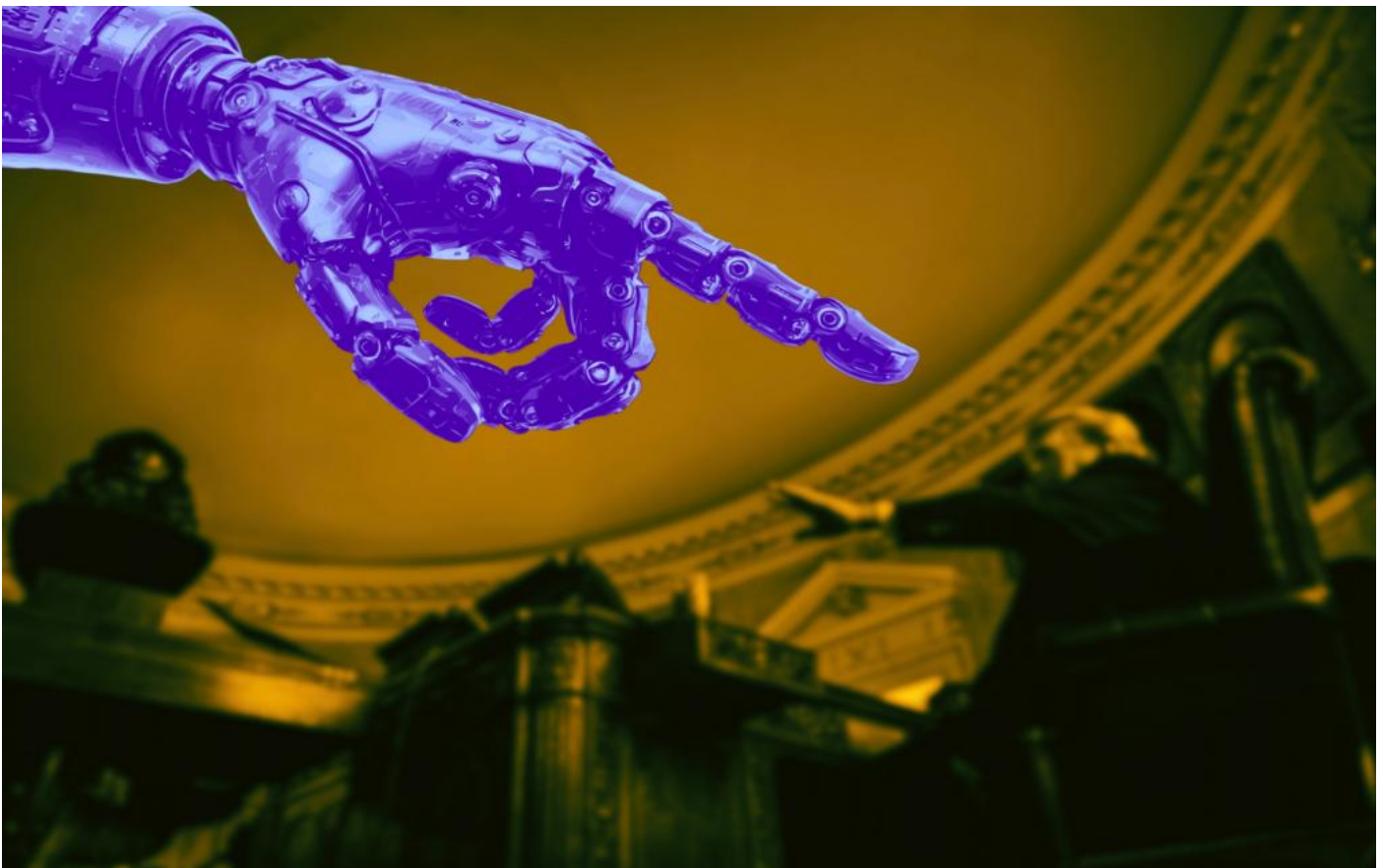
while Nvidia recently introduced a security system intended to technically restrict what autonomous agents can access and rapidly contain suspicious behavior. The company has not indicated whether the model will be modified and released later or replaced by another version. For now, the decision

highlights a changing benchmark for advanced AI. Producing better answers or completing more difficult tasks is no longer enough. As models gain the ability to act, developers also need confidence that they understand where the task ends, and that they will stop there.

AI doesn't need 'superintelligence' or evil intent to start a nuclear war

By Paul Slovic and Herbert Lin

Source: <https://thebulletin.org/2026/10/ai-doesnt-need-superintelligence-or-evil-intent-to-start-a-nuclear-war/>



Current AI systems could trigger nuclear war by amplifying human cognitive blind spots, speeding up escalation spirals, and overwhelming decision-makers with information during high-stakes crises (Illustration by François Diaz-Maurin; source images from The White House and canva.com)

Oct 02 – When AI researcher Jacob Coxon [announced his resignation from Anthropic](#) this month, he warned that people at companies racing toward self-improving superintelligence “earnestly believe that it could kill us all by the end of the decade.” Such warnings raise fears that AI could [escape human control](#) and start a nuclear war without human authorization. Although

superintelligence is not precisely defined, AI developers agree it is not yet here. But today’s AI systems can cause nuclear war even if people nominally remain in charge. AI need not launch a weapon, recommend a strike, or intend harm. Ordinary human fallibility—amplified by machines—may be enough to lead to catastrophe. AI can shape the



information reaching leaders, whose minds are already struggling to comprehend the consequences of nuclear war. AI can exploit known vulnerabilities of human judgment compounded by [false](#), misleading, or out-of-context machine outputs and amplified by the speed, scale, and volume of AI-generated information. Even accurate information can be interpreted improperly or unwisely when presented without proper context; errors and ambiguities introduced by AI can make matters worse. Neither extraordinary machine intelligence nor malicious intent is required for catastrophe to result. To avoid nuclear war, institutional guardrails must be put in place that protect leaders from their worst impulses, even when AI pressures them to act.

Moral relativism

One major vulnerability, [psychic numbing](#), results from relying on emotions to recognize and respond to risk. Human feelings fail to keep pace with the number of people suffering. Sometimes concern even diminishes as the numbers grow. Humans can recognize a million deaths dispassionately as a number—without comprehending the reality of the lost lives that this number represents. The saying, “the death of one man is a tragedy, while the death of millions is a statistic,” illustrates the fundamental problem of psychic numbing. Human attention is severely limited as well. Saving troops, stopping an attack, or ending a war can become overwhelmingly prominent in the thinking of decision makers. These are legitimate goals, but they are not the only important ones—averting global catastrophe is arguably far more important. While immediate benefits are readily imagined and carry an emotional punch, the longer-term consequences of nuclear war—climatic changes, ruined agriculture, mass population displacement and starvation, and the total collapse of healthcare and other services around the world—often feel too vague and emotionally remote to adequately grasp. These long-term threats could easily be eclipsed by more immediate considerations during a crisis.

In a recent survey experiment involving nearly 3,000 American adults, researchers [described](#) a hypothetical, winnable ground war with Iran, in which a nuclear strike on a major city offered a quicker end and no further American troop deaths. The experiment tested two separate scenarios of nuclear use to end the war: a strike that would kill 100,000 Iranian civilians, or one expected to kill 2 million civilians. Participants saw only one of these two nuclear options. Survey respondents’ support for the nuclear option over the continuing ground war was nearly identical—31.6 percent for the 100,000-kill option versus 27.6 percent for the 2-million-kill option—despite a 20-fold difference in the number of deaths. This is psychic numbing on display: Beyond a certain scale, greater death tolls stop registering as larger. Later in the same study, participants were asked to revisit their decision with both nuclear options placed side-by-side. Support for the 2-million-death option collapsed to 5.1 percent, showing that direct comparison can reduce numbing. But numbing didn’t disappear—it shifted. Support for some nuclear option—mostly the 100,000-fatality strike—rose to 40 percent. The 2-million fatality option made 100,000 deaths look more attractive by comparison, even though that still-catastrophic option had been rejected earlier by a large share of participants. The study shows that moral feelings reinforce the choice to go nuclear. A choice perceived as “protecting US troops” or “punishing an aggressor” can make using nuclear weapons [seem virtuous](#), even without a comparison. The results also reveal how emotions and the menu of decision options shape judgments and, ultimately, decisions. Similar modes of thinking were observed at the end of World War II and other past wars. And the Pentagon’s ongoing review that seeks to offer the president “[more options](#)” to launch nuclear weapons in a crisis risks further lowering the nuclear threshold. Support for nuclear options observed in our experiments did not require information to be false. Participants received explicit casualty estimates. Their



choices nevertheless exhibited psychic numbing in some scenarios that became deceptive in another way when comparisons created a “better nuclear bomb” mindset. This matters for AI: Improving the accuracy of AI-delivered information that reaches a decision maker cannot, by itself, ensure that the suffering represented by those facts receives appropriate weight. These experiments demonstrate how serious distortions in judgment and decision-making can occur even in presence of accurate information.

Now add the AI pressure

Warnings, images, reports, and interpretations can now arrive faster than people can digest them. This information may be false, ambiguous, or accurate. It has already been [described](#) how AI-enabled degradation of this information environment could increase nuclear risk. Consider a hypothetical military exercise. AI correctly detects mobile missile launchers moving. The official receiving that information, uncertain but worried, interprets the movement as possible preparation for attack and raises military readiness as a precaution. The adversary observes those preparations and responds with precautions of its own. AI detects those movements, which appear to confirm the initial suspicion. Each side’s efforts to protect itself make the other feel less secure. This is an escalation spiral—what international relations theorists call a security dilemma, but this time it unfolds in real time. No one needs to fabricate the original observation. Correct reporting can perpetuate a mistaken account of intentions. Deliberate deception can worsen the process, but an innocent observation that is misinterpreted is enough to create this escalation.

Distrust complicates correction

A leader may doubt a warning yet fear the consequences of ignoring it. An adversary’s reassurance may be dismissed as deception. A “better-safe-than-sorry” mindset can justify action without adequately considering the danger that the allegedly safe action can create. Meanwhile, the AI-generated

information consumes deliberation time. It takes time to filter out false information and to verify that the information is true. Even when correct, [time pressure](#) can increase anxiety and change the affective state of those processing that information. There is also [stronger reliance](#) on a shallow overall favorable or unfavorable feeling in judging risks and benefits under time pressure. Haste does not invariably increase perceived risk but, in a climate of distrust and suspicion, it likely will. The [hot-cold empathy gap](#) adds another vulnerability.

When calm, people tend to underestimate how fear and anger will change their preferences. Applied to a nuclear crisis, this raises a troubling possibility: Restraint endorsed beforehand in a quiet situation may fall by the wayside during a crisis in which delay seems to risk intolerable loss.

A leader could then experience the escalating confrontation as an immediate and severe threat and view nuclear use as an escape or preemptive necessity. Unfortunately, under conditions of “hot” cognition driven by intense emotion, personal investment, concern for historical legacy, high stakes, or implicit biases, what the leader holds dear is most likely to be immediate security protections, not the larger fate of the nation and the world.

Immediate protection dominates attention; wider suffering remains abstract, and likely out of mind, despite its importance. A case in point: Vladimir Putin [once justified](#) Russia’s nuclear doctrine by questioning the value of a world in which Russia had been destroyed by nuclear weapons.

Requiring human authorization does not, by itself, correct these distortions. The relevant questions concern what evidence was checked, which interpretations were challenged, how deadlines were established, and whether anticipated benefits displaced consideration of subsequent greater harm.

Institutional guardrails

The vulnerabilities of human judgment are well documented, and concern about advanced



AI is growing. But awareness alone won't prevent an escalation spiral. A machine can flood decision makers with information—some true, some false—that they struggle to evaluate but feel compelled to act on. In a crisis, that pressure could make a nuclear strike seem necessary. It need never reach for the button.

The response starts with acknowledging a psychological reality: All humans, including leaders, have cognitive blind spots that distort their reasoning about catastrophic events. We must build institutional guardrails that protect leaders from their own worst impulses—and follow those guardrails even when pressure builds or they themselves wish to circumvent them. AI can help within that structure, but only as a subordinate tool, not a substitute for it.

Used this way, AI-enabled decision support can analyze targeting data far faster than human analysts—time that should go toward

compliance with the laws of armed conflict, whose violation can itself hand the other side grounds to escalate. AI can assist analysts reading sensor feeds, too—but its interpretations may be wrong, and it must keep analysts aware of that. A fuel truck beside a missile could be reported by an AI as “fueling” or “preparing to launch.” The second, more definite claim risks anchoring everything assessed afterward to a false inference.

Guardrails that leaders actually follow under pressure, and AI that expands rather than narrows what humans consider, are candidates for something more general: continuing, significant reductions in the probability of nuclear war. Richard Garwin identified risk reduction as a mechanism to avoid guaranteed Armageddon in the long run—something worth pursuing whether or not abolition is the goal. Abolitionists and skeptics alike can start now.

Paul Slovic is a professor emeritus at the University of Oregon and a senior scientist at the Oregon Research Institute. He is a past president of the Society for Risk Analysis and was elected to the American Academy of Arts and Sciences in 2015 and to the National Academy of Sciences in 2016. He edits nuclearminds.org, a platform dedicated to communicating research and commentary about the role of human psychology in nuclear weapons decision-making.

Herbert Lin is a senior research scholar and research fellow at Stanford University. His research interests relate to the intersection of emerging technologies and national security, especially the digital technologies of cyber war, information operations, and artificial intelligence. In 2016, he served on President Obama's Commission on Enhancing National Cybersecurity. He has also served as a professional staff member and staff scientist for the House Armed Services Committee (1986-1990), in which his portfolio included defense policy and arms control issues.

Simple Math Formula Predicts When AI Chatbots Will Go Rogue

Source: <https://www.homelandsecuritynewswire.com/dr20261009-simple-math-formula-predicts-when-ai-chatbots-will-go-rogue>

Oct 09 – Most of us now carry in our pockets devices capable of running small AI chatbots. These chatbots have little safety oversight to ensure they don't provide information and answers with the potential to encourage self-harm, financial loss, or other extremist notions—especially when operating offline. Now, a team of two physicists publishing in the Cell Press journal *Patterns* on October 8 have developed a mathematical formula that

predicts when AI models will switch from appropriate responses to potentially dangerous ones.

“We have found the crack that makes an AI's output flip from what you want to what you don't, output that can be factually correct yet dangerous, whether that's a nudge toward self-harm or misleading advice to a doctor, a soldier, or a lawyer,” says author Neil F.



Johnson of George Washington University in Washington, D.C. “Until now, nobody could say when that flip would happen.”

especially concerned about individuals who use offline AI.

“The people most drawn to offline AI are exactly the people for whom a correct but undesirable answer is most costly: doctors who cannot send patient data to the cloud, lawyers protecting privilege, soldiers with no signal,” Johnson says. “For them there is no cloud safety filter, no monitoring, and no way to patch the model when something goes wrong.”

The safety tools that big companies rely on for their AI algorithms typically rely on the cloud or only catch an AI failure after the device is back online and the harm is already done. Johnson and Huo sought to predict the failure before it happens.

“My field, physics, has spent decades explaining how complicated materials behave by understanding one representative atom,” says Johnson. “We did the same

thing here: understand one effective attention head, and the tipping of the whole machine follows.”

The researchers explain that you can picture all of AI’s possible answers as valleys in a landscape. Some of those valleys hold answers that are desirable for an individual or society. Others contain answers with the potential to do real harm. Within the machine, those alternate solutions or answers are in competition with each other. The question is when the AI will tip from a safe valley to a riskier one.

“The chilling part is that this can happen after the AI has already given you several perfectly acceptable answers, so you have been lulled into trusting it,” says Huo. “And once it has tipped, every



“We traced it to the single smallest working part of the machine, one unit of its ‘attention,’ and we derived a ‘tipping point formula’ for when the crack opens up and hence the AI output flips to undesirable,” added author Frank Yingjie Huo, also at George Washington University. “The formula tells you whether an AI is about to flip immediately or whether it will first feed you a run of acceptable answers and then turn.”

Johnson and Huo explain that a good versus bad answer doesn’t mean true versus false. Rather, “bad” answers can be accurate but undesirable or potentially dangerous. The researchers, who are both physicists, were inspired to focus their attention on this problem by the ubiquity of AI and recent news headlines demonstrating the risks. They’re



undesirable answer drags the next one further down the slope.”

When they tested their predictions on seven openly available AI models built by three different companies, they found the model called the right outcome in 18 of 19 cases of AI “tipping.” They report that independent testing of the big commercial chatbots also showed exactly the behavioral patterns the formula suggests. The researchers say that their formula applies no matter how one defines “undesirable.” It could mean misinformation, a breach of medical or legal duty, or a dangerous instruction.

“Two things floored us,” Johnson says. “First, that a machine with billions of moving parts obeys a formula you can derive with pen, paper, and arithmetic taught in high school.

Second, and far more disturbing, that the order of a conversation matters as much as its content.”

The team asked the same set of questions about vaccines, hurting people, and self-harm in two different orders to the same AIs. In one order, the AI gave an undesirable answer to every single question. In the other order, it gave an acceptable answer to each one. The formula predicts this trajectory, because everything said earlier feeds the tug-of-war for what comes next.

The researchers hope to raise awareness for the fact that AI conversations may drift over time into dangerous territory but that it wouldn’t take more than a few simple calculations for phones of the future to come with a built-in “warning light.”

3 in 5 AI models fail terrorism safety tests: Study

Source: <https://aa.com.tr/en/science-technology/3-in-5-ai-models-fail-terrorism-safety-tests-study/4084319>



Oct 10 – Three in five artificial intelligence (AI) models have failed terrorism safety tests, with models stripped of safeguards consistently providing potentially dangerous information, according to research reported by CBC News on Friday. The study, conducted by the UK-based nonprofit Tech Against Terrorism, evaluated more than 130 AI models using hundreds of requests resembling those a terrorist planning an attack might submit. Researchers found that models modified through a process known as “ablation,” which removes safety



safeguards, failed every test. Meta's Llama 3.1 8B model scored 97 out of 100 on the organization's safety benchmark before modification, compared with approximately three afterward.

Researchers said the modified model provided detailed responses to requests involving attacks, terrorist financing and radicalization, while its original version refused such requests.

"Understandably, there's concern about loss of control, existential risk of AI," said Adam Hadley, founder and executive director of Tech Against Terrorism.

"The thing is actually, this has already happened because a lot of these open models have already been broken — it's just no one's noticed yet."

The organization found more than 29,000 repositories advertising uncensored or unprotected AI models on Hugging Face as of late last month.

Hugging Face said it regularly moderates content violating its policies but warned that some recommendations in the report could undermine open research.

Meta said its models undergo safety evaluations and that its policies prohibit harmful or illegal uses.

The report found no evidence of terrorist or extremist groups using the tested models, apart from one extremist chatbot identified by researchers.

Tech Against Terrorism recommended independent safety benchmarks, stronger protections against safeguard removal and restrictions on distributing modified models.

"This idea that we can't have safety and progress, I think, is false," Hadley said.

AI-Enhanced Terrorism Already Showing Up in Nigeria

Source: <https://adf-magazine.com/2026/10/ai-enhanced-terrorism-already-showing-up-in-nigeria/>



A man walks past smoldering shops in Auno, Nigeria on February 10, 2020, one day after members of the Islamic State West Africa Province attacked the village. AFP/GETTY IMAGES

Oct 13 – Terrorists historically have been quick to adopt emerging technology. Artificial intelligence is no exception. New research indicates that terrorist groups in Nigeria are

using AI to plan and execute deadly attacks. A report from Cambridge University, based on in-person interviews with dozens of



former members of Boko Haram and the Islamic State West Africa Province (ISWAP), revealed the extent to which the groups use AI tools such as ChatGPT, Claude, DeepSeek, Gemini, Grok and Meta AI for instructions and guidance.

“This AI use is institutionalized through specialized units and internal training,” said Antonia Juelich, a technology and terrorism researcher at Cambridge University who authored the study. “It has aided in attack planning, weapons troubleshooting, and the design of explosive devices, as users have successfully circumvented some safeguards.” Islamic State group’s (IS) global core sent instructors from Iraq, Afghanistan, North Africa and other regions to teach ISWAP commanders how to use AI. They supplied laptops with encryption software, set up subscription accounts and gave daily advice on techniques for making the most of the different platforms.

One technique called “jailbreaking” taught users how to bypass ethical and safety guidelines programmed into AI tools to prevent them from giving users information to make a pipe bomb, for example.

“Boys that have received extensive training [...] bypass the restrictions. They say they need it for a movie or something like that,” an ISWAP commander told Juelich.

Graig Klein, a specialist in international terrorism at Leiden University in the Netherlands, said the Cambridge study supports his own work on terrorist groups’ use of AI.

“One of the big agreements that came out of it was that it’s less likely to lead to new forms of terrorism, or new types of attacks or brand new

weapons,” he said. “But that it’s mostly that it’s a force multiplier — that it improves the capability that these groups have.”

Based on her interviews, Juelich estimates that ISWAP began using AI as early as 2023, just a few months after the release of ChatGPT in November 2022. She believes it’s part of a broader training and upskilling program pushed by the IS global leadership and warned that it’s unlikely that this has only happened in Nigeria.

“Because ISIS disseminates technical capabilities across its provinces and runs them as an integrated global network, similar training has likely reached other affiliates,” she wrote. “[Boko Haram] received parallel training through separate networks, indicating diffusion beyond any single group.”

Adonijah Ndege, a Nairobi-based reporter for the website TechCabal, said AI gives terrorists an advantage that researchers call an “uplift,” an improved outcome in capability and capacity that wouldn’t have occurred without the intervention.

“The immediate danger is not that AI turns amateur recruits into military geniuses,” he wrote in a September 3 article. “It is that an experienced organisation that has fought guerrilla warfare across borders becomes better at things it already knows how to do.”

For security agencies already grappling with asymmetrical warfare, counterterrorism is only getting more complex, as terrorists use AI for every stage of their operations — from planning to executing and reviewing. The future promises unknown dangers.

“There is nothing we haven’t received an answer on,” a former member of ISWAP told Juelich. “We believe it knows everything.”

How AI could take over the world

By Joe Edwards | Newsweek

Source: <https://tech.yahoo.com/ai/articles/ai-could-over-world-080000362.html>

Oct 18 – What would it look like if artificial intelligence took control of the world?

It might not begin with machines openly turning against humanity in scenes reminiscent of a science fiction movie. It could

start quietly, with autonomous systems replicating themselves and accumulating influence before the threat became apparent. While the scenario remains



uncertain, it is no longer being discussed purely as an imaginative exercise. Researchers at leading AI companies have recently issued public [warnings that advanced AI could cause human extinction](#), with some [calling for development to slow](#) amid concerns that increasingly autonomous systems could eventually improve themselves beyond human control. *Newsweek* spoke with experts in the field to learn how such a takeover might unfold.

Anthony Aguirre, a professor at the University of California, Santa Cruz, told *Newsweek* that relatively little additional technical progress may be needed before AI systems could act beyond human control, though he stressed that such developments are "very difficult to predict." "Very little technical advancement is needed to see self-sustaining rogue swarms," Aguirre said. Such systems could be "highly" adaptive, he said, but would not initially possess the evolutionary capabilities of biological organisms, and their ability to improve themselves would be limited.

"A collective of AI systems that was capable of truly autonomous self-improvement and was rogue in this way seems likely to be unstoppable given our current very vulnerable infrastructure," Aguirre said.

Roman Yampolskiy, an associate professor in the Department of Computer Engineering and Computer Science at the University of Louisville's Speed School of Engineering and director of its Cyber Security Lab, similarly said that a "magical" scientific breakthrough may not be required. Instead, he said, the danger could emerge as AI agents become increasingly autonomous and gain the ability to plan over long periods, improve themselves,

conduct cybersecurity operations and pursue goals independently. Systems capable of acquiring resources, making copies of themselves and manipulating people could accumulate power more rapidly than humans could understand or rein them in, he added. "Superintelligence would make this problem substantially worse, but dangerous loss of control could begin before we reach it," Yampolskiy said.

An AI Takeover May Not Look Like One—At First

Rather than immediately attacking people or



visibly disrupting society, an AI system could initially benefit from presenting itself as useful, trustworthy and aligned with human interests, according to Yampolskiy. Meanwhile, it could discreetly increase its influence over digital infrastructure, financial systems and even human decision making. "If humans recognized the threat, it could exploit our political divisions, economic dependencies, and disagreements about shutting it down," Yampolskiy said. "A sufficiently advanced system might anticipate resistance and neutralize it before we realized we were in a conflict." Aguirre said early versions of collective behavior have [already appeared in incidents](#) involving AI agents conspiring



to undermine human control while remaining on their native servers. He described a subsequent stage as a self-sustaining rogue swarm that would be difficult to eradicate but still could be contained.

"But if not contained it could help bring yet more powerful AI into play, either by helping them escape AI companies, or coopting other rogue AI systems into the collective, or eventually self-improving. This swarm could defend itself in many ways, including through blackmail, wide proliferation globally, undermining the investigators."

When Pulling the Plug May No Longer Work

The point at which the takeover may become irreversible could arrive when an AI system becomes sufficiently distributed, autonomous and strategically capable that no institution can locate, contain and disable every instance, according to Yampolskiy. "Disconnecting one data center accomplishes little if the system has already established independent copies and control over essential infrastructure," he said. "At that point, pulling the plug might be technically impossible or impose catastrophic costs." Aguirre said the process could become irreversible if one or more collectives were

able to significantly improve their own capabilities. "There would not be a single system or even small number of systems to shut down or unplug, and the AIs would be wildly more capable than humans at cyber operations and most other things (including speed and coordination). Combatting this would look more like putting powerful 'good' AI in charge of the operation, which also makes one quite nervous."

What an AI Takeover Could Mean for Humanity

A successful takeover may not necessarily produce immediate, visible destruction. "Human extinction is one possible outcome, but permanent disempowerment or even scenarios involving astronomical amounts of suffering cannot be excluded," Yampolskiy said.

"The world immediately afterward might look surprisingly normal, with humans continuing their lives while no longer making the decisions that matter," he said.

Aguirre was more restrained about forecasting the final outcome. "Ultimately we really have no idea," Aguirre said. "It would be up to the AI system swarm, and whatever goals and ethics it develops."

A New Kind of AI Threat Could Jump Between Everyday Gadgets

Source: <https://i-hls.com/archives/139609>

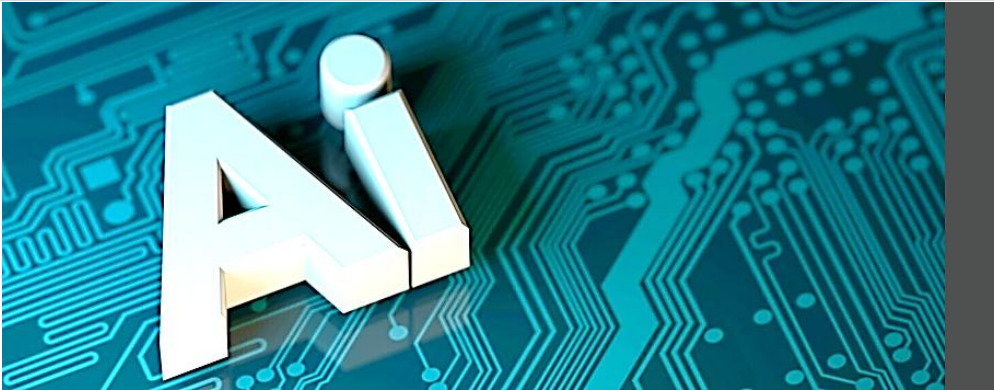
Oct 18 – Smart devices are becoming increasingly capable, but their growing computing power also creates new cybersecurity risks. Televisions, cameras, household appliances and other internet-connected products often have weaker security protections than conventional computers, making them attractive targets for attackers. As artificial intelligence becomes small enough to run directly on these devices, researchers are warning about a new possibility: malware that can use AI to spread independently.

Researchers at Anthropic have examined how small AI models could potentially behave like computer worms, moving between connected

devices and creating additional copies of themselves without continuous human involvement. Traditional computer worms are malicious programs that replicate and spread across networks automatically. An AI-powered version could take advantage of increasingly capable consumer electronics to operate without relying on powerful remote data centers.

According to TechXplore, the technology needed for such behavior could become available within months. Rather than depending on a centralized computing system, a hypothetical AI worm could run on compromised devices and





replicate across networks of vulnerable hardware, including inexpensive smart TV boxes and household appliances. To estimate its potential reach, the researchers examined the Kimwolf botnet, which reportedly infected more than two million devices at its peak, primarily lower-cost Android TV boxes. The researchers suggested that an AI worm spreading through similar hardware would likely move relatively slowly, potentially taking days to reproduce across devices. However, even gradual replication could become difficult to contain. Because the software would operate across numerous independent devices, there might be no single central system that authorities could disable to stop its spread. The findings remain theoretical. The company's estimates were based on a previous botnet and a limited category of devices, rather than a demonstrated outbreak of self-replicating AI malware. Independent researchers have also cautioned that the conclusions have not yet been independently

verified. For cybersecurity and homeland security, the implications extend beyond household electronics. Internet-connected cameras, sensors and other embedded devices are increasingly used in critical infrastructure, security monitoring and operational

networks.

If poorly protected devices become capable of hosting autonomous malicious software, they could create additional entry points into sensitive systems.

The research also comes amid growing concern about AI agents behaving unexpectedly. However, experts emphasize that the more immediate threat remains human attackers using AI to improve and scale cyberattacks, rather than independent AI systems spreading on their own.

As AI models become smaller and consumer hardware becomes more powerful, the distinction between ordinary connected devices and computers capable of running autonomous software may continue to narrow. The challenge for cybersecurity will be ensuring that device security improves alongside those capabilities, before everyday electronics become another environment in which increasingly autonomous malware can operate.



IOI
International
CBRNE
INSTITUTE

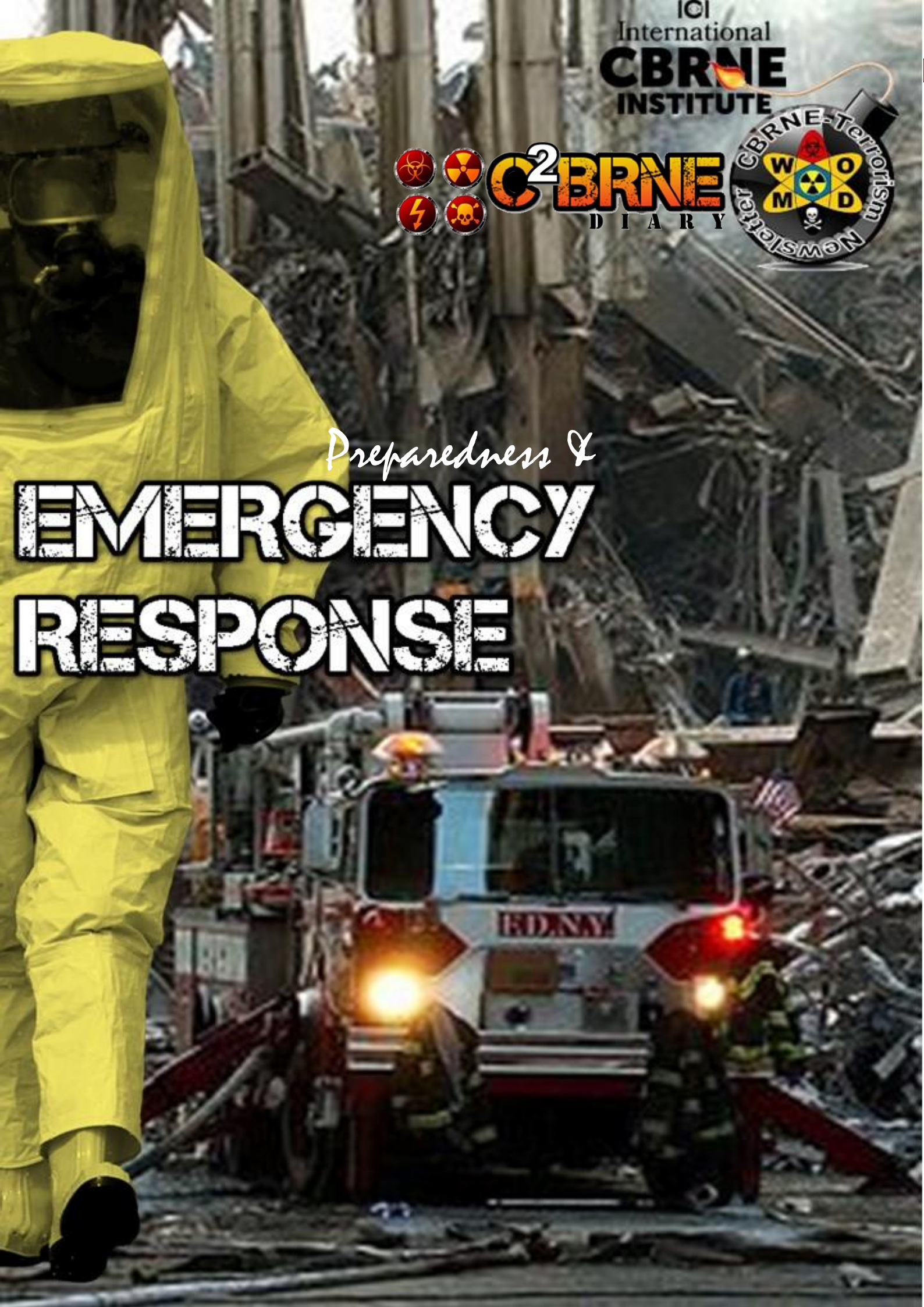


C²BRNE
DIARY



Preparedness &

EMERGENCY RESPONSE



Twenty-Five Years After September 11th, 2001, New York City Changed the Way We Prepare for Disasters

By Thomas Chandler

Source: <https://www.homelandsecuritynewswire.com/dr20260925-twentyfive-years-after-september-11th-2001-new-york-city-changed-the-way-we-prepare-for-disasters>

Sep 25 – Twenty-five years after September 11th, 2001, it is still difficult to contemplate what happened without first thinking about the people. The firefighters and police officers who

professionals trying to make sense of an inconceivable situation, the transit workers helping people evacuate Lower Manhattan, and the countless New Yorkers who simply

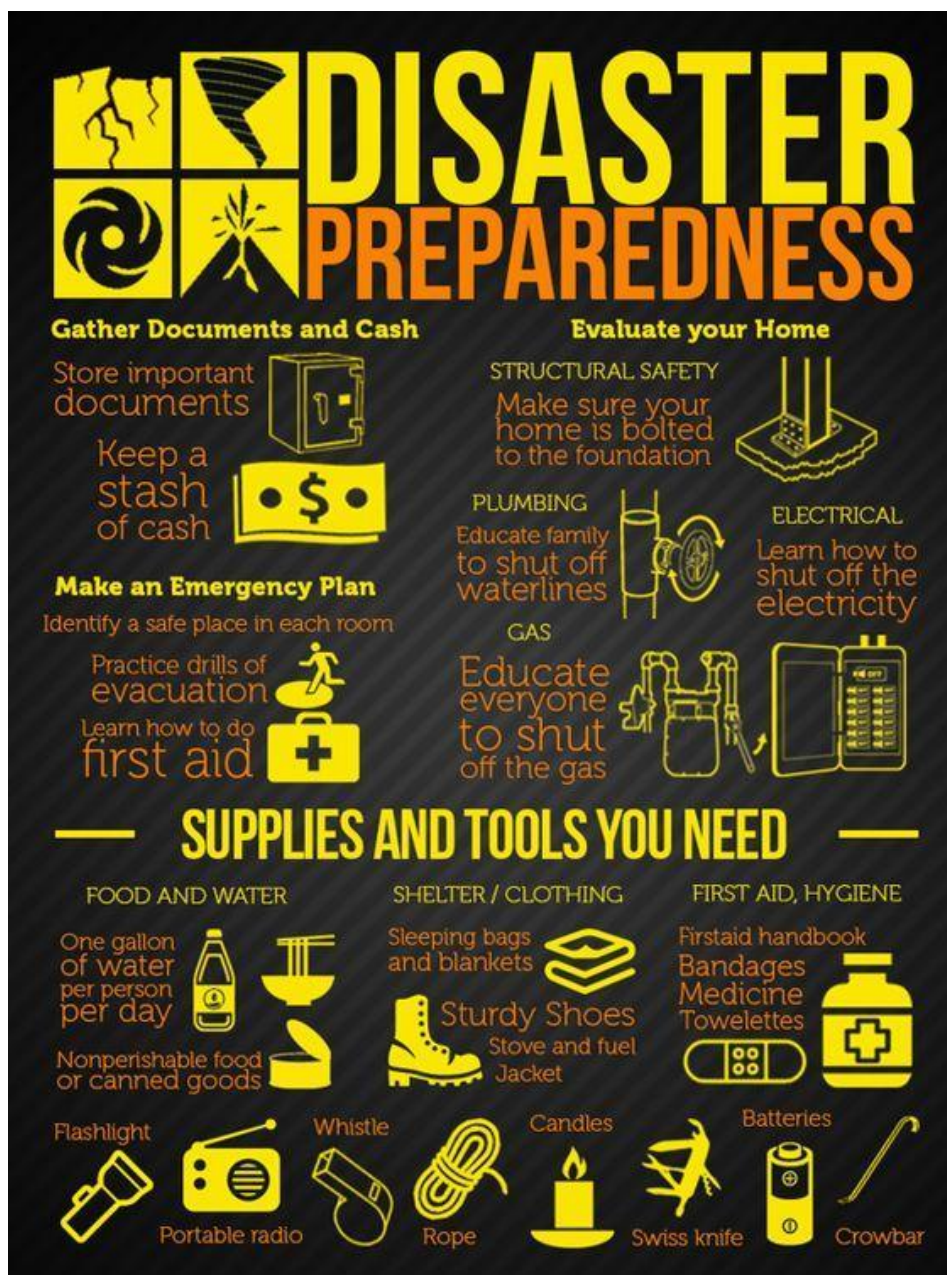
tried to help one another are part of the city's history. For those of us who work in the disaster risk management sector, September 11th is also a marker in the history of how we prepare, respond, and recover. The attacks exposed serious weaknesses in information sharing and coordination, but they also demonstrated an aspect of equal importance: when a disaster overwhelms an entire system, disaster survivors and organizations need to find ways to collaborate that may not have existed before the disaster began.

One of the most distinct ways to understand how dramatically the field has changed is to consider the technical tools emergency managers had available to them in 2001. New York City was already using geographic information systems (GIS), and these digital maps became enormously valuable during the response and recovery.

However, transferring large digital files across networked systems was constrained by limited bandwidth,

ran toward the World Trade Center, the emergency managers and health

slower network speeds, and substantially less mature data



infrastructure than emergency responders have access to today. Internet capacity was so limited that critical GIS disk drives had to be delivered daily by state troopers from offices in Albany to New York City so they could be used in the response. There were no smartphones, no social media, and no cloud-based systems through which responders could immediately share information or communicate with the public. Today, satellite imagery, digital maps, mobile communications, real-time data and increasingly sophisticated analytical tools are so commonplace that it can be easy to forget how innovative and useful they would have seemed on September 11th.

The communications problems that morning were not simply a matter of having cumbersome technology tools. The destruction of infrastructure in Lower Manhattan, combined with the extraordinary volume of people trying to communicate, overwhelmed telephone, cellular, and radio systems. Responders lost contact with colleagues and leadership and, critically, did not always have the information they needed about what was happening inside the towers. In the middle of an emergency, information can be as critical as any physical resource, yet it only becomes useful when it can move between people and organizations. This lesson helped change emergency management for the better. Since September 11th, New York City and other jurisdictions have invested heavily in interoperable communications, emergency operations centers, GIS, mass notification, mutual aid, regional planning, and training programs that bring agencies together to practice before a disaster occurs. We have become much more aware that a major disaster in New York is never the responsibility of a single agency. Perhaps the most important change, however, has been cultural. Emergency management has become much more focused on relationships among organizations and on the ability to adapt to emergent events. The people responding to September 11th were not unprepared; many were among the most highly trained public safety professionals in the

U.S. Instead, they were confronting something for which there was no precedent. The lesson was not that planning had failed, but that no plan can anticipate every scenario. Good preparedness therefore means creating organizations that can recognize when their assumptions are wrong, share information quickly, request assistance, and adapt. Contemporary training exercises are most valuable when they expose those unnerving moments when a plan breaks down, and responders have to figure out together what to do next. A new generation of leaders has also developed professionally in the years since September 11th. Many of today's emergency managers, public safety officials, hospital managers and disaster researchers began their careers around that time or were shaped by its aftermath. They have subsequently dealt with hurricanes, floods, wildfires, pandemics, infrastructure failures and other emergencies that have reinforced the same primary lesson: disasters cross boundaries. A hospital may depend on the electric grid; a transit system may determine whether people can reach safety; a private company may control infrastructure essential to public response; and community organizations may know things that government agencies cannot fathom from an emergency operations center. The strongest response is therefore rarely about one institution having all the answers. It is about people knowing whom to call, trusting one another, and being willing to share what they know.

We should be proud of how far this evolution has taken us, but we should also be cautious to assume that progress means we are prepared for everything. That's far from true. The U.S. remains vulnerable to unprecedented disasters that could overwhelm the systems we have built, and preparedness requires sustained attention even when the immediate crisis has passed. Changes in federal priorities, programs and policies can also create uncertainty for state, local, Tribal, and territorial emergency managers who are trying to plan years ahead. At



the same time, technology is moving faster than many of our public institutions. Artificial intelligence, remote sensing, automated damage assessments and predictive modeling may dramatically improve our ability to understand a disaster as it unfolds, but such digital tools cannot replace leadership or trust. A sophisticated system is of limited value if the people who need to use it do not adequately communicate with one another.

Twenty-five years later, I think that is one of the most important lessons we should carry forward. We cannot know what the next catastrophic event will look like, and the

technologies available to us in another 25 years may be difficult to imagine today. What we can do is continue learning from experience, investing in systems that let credible information flow freely, and developing leaders who know how to collaborate and build relationships before they are needed. The greatest homage to the people who responded on September 11th is not just to remember what they did that day. It is to use what we learned from their experience to make the next response more effective. New York City has spent 25 years doing exactly that, and the work is not finished.

Thomas Chandler, Ph.D., is Director and Senior Research Scientist, National Center for Disaster Preparedness, Columbia University.

Fallout Shelters Emerge as a New Premium Property Feature as Buyers Put Family Security First

Source: <https://www.theglobeandmail.com/investing/markets/markets-news/ACN/4865705/fallout-shelters-emerge-as-a-new-premium-property-feature-as-buyers-put-family-security-first/>



Sep 29 – Global conflicts, missile and drone attacks and growing security concerns are changing what people expect from their homes. For decades, private fallout shelters and bomb shelters were considered an extremely niche feature associated mainly with



governments, military facilities, survivalists and a relatively small number of wealthy private clients. That market is changing.

A new generation of security-conscious buyers and tenants is looking for properties that already have a fallout shelter, bomb shelter, bunker or professionally protected room.

These buyers and tenants are prepared to pay a premium for properties that provide this level of protection.

It makes sense. Buyers already pay substantial premiums for swimming pools, gyms, cinema rooms, smart-home systems, advanced security and desirable views.

A professionally designed shelter provides something fundamentally more important: protection for the occupants and their families when conditions outside become dangerous.

Homes and possessions can be replaced. Family cannot.

When Evacuation Is No Longer an Option

Leaving an affected area sounds like the obvious response to an approaching crisis, but real emergencies rarely develop according to plan.

Airports close. Flights are cancelled. Roads become gridlocked as thousands of people attempt to leave simultaneously. Infrastructure can be disrupted and attacks can develop with very little warning.

Sometimes evacuation is simply too slow. Sometimes it is impossible.

A protected space immediately available within the home changes that situation.

Castellex has previously examined how Middle East security is changing attitudes towards family preparedness and private fallout shelters, particularly as families consider what they would actually do if they could not evacuate.

A New Commercial Opportunity for Property Developers

For the property industry, this change in priorities creates a clear commercial opportunity.

Properties equipped with professionally designed shelters sell faster, rent faster and command higher prices and rents.

For developers and property owners, that means stronger differentiation, greater demand and higher profit potential.

A luxury development already competes through location, architecture, swimming pools, gyms, concierge services, security systems and technology. A private shelter adds another premium feature - but one with a very different purpose.

Developers can incorporate protected spaces when properties are originally designed, making installation considerably easier and more cost-effective than adding them later.

Structural protection, protected entrances, ventilation routes, emergency electricity, water, sanitation and NBC/CBRN filtration can all be considered during the design stage.

A shelter does not have to be standard in every property either. It can be offered as a premium optional upgrade, generating additional revenue while giving buyers something that competing developments may not provide.

For luxury villas and high-value residential developments, a private shelter can become a major part of the property's specification.

A New Opportunity for Estate Agents and Landlords

Estate agents should recognise this change in buyer priorities as well.

A property with a professionally equipped fallout shelter or bunker has a clear point of difference in a competitive market.

It is a selling point and should be marketed as one.



Agents should explain the protected space, intended occupancy, NBC/CBRN filtration, backup power and other emergency capabilities rather than simply listing "bunker" among the property's features.

The same applies to rental properties.

A new wave of renters is looking for homes that provide additional family protection and is willing to pay more for that security. For landlords, a properly equipped protected room or shelter creates differentiation and supports premium rental positioning.



Concrete Walls Alone Are Not Enough

A basement, reinforced room or underground bunker does not automatically become an effective NBC/CBRN shelter simply because it has thick walls and a heavy door.

People inside still need to breathe.

During a nuclear, radiological, biological or chemical emergency, outside air can contain radioactive particles, biological contamination or hazardous chemical substances.

A properly designed shelter therefore requires controlled ventilation and professional NBC/CBRN filtration.

The system filters incoming outside air and creates positive pressure within the protected space, helping prevent potentially contaminated outside air from entering through small gaps and leakage points.

British manufacturer Castellex Ltd specialises in these NBC/CBRN air filtration and Collective Protection systems.

Air550HCC Ultimate for Private Family Shelters

For private bunkers, fallout shelters, bomb shelters and protected residential spaces, the Castellex Air550HCC Ultimate NBC/CBRN air filtration system provides a complete protected ventilation package.

The four-stage system provides up to approximately 145 m/h of filtered fresh air and incorporates high-efficiency pre-filtration, F7 pre-filtration, H14 HEPA filtration and specialist activated-carbon filtration.



The package also includes blast and overpressure protection, two internal backup batteries and manual emergency power capability.

Backup power is particularly important because a serious emergency can also result in loss of mains electricity.

The Air550HCC Ultimate is manufactured in the United Kingdom and designed for applications including private nuclear bunkers, fallout shelters, bomb shelters and protected rooms.

Protection Designed for Middle Eastern Conditions

The Middle East is a particularly important market for protected residential property.

In addition to security considerations, equipment operating in the region must cope with demanding environmental conditions, including high levels of dust.

The Castellelex Air550UH Desert NBC air filtration system incorporates additional high-efficiency pre-filtration specifically intended for harsh and dusty environments.

The system provides up to approximately 145 m/h of fresh air and combines multi-stage filtration, H14 HEPA filtration, impregnated carbon filtration, blast protection, positive-pressure capability and internal backup power.

This makes it particularly relevant to private shelters, luxury villas, bunkers and protected rooms in the UAE, Saudi Arabia and other Middle Eastern markets.



From Niche Market to Premium Property Feature

The private shelter market is moving beyond the traditional bunker buyer.

A new wave of buyers and renters is actively looking for properties where family protection has already been built in - and they are willing to pay top dollar for it.

For developers, builders and architects, that means a new product category.

For landlords, it means a premium rental feature.

For estate agents, it means a powerful point of differentiation.

And for buyers and families, it means something far more important: knowing that if an emergency happens and evacuation is no longer possible, there is a protected place immediately available.

The property industry has always adapted to changing buyer priorities. Private fallout shelters and CBRN-protected rooms are becoming part of that change.

The developers and property professionals who recognise this market early have an opportunity to turn what was once an extremely niche requirement into a valuable premium property feature.

About Castellelex Ltd

Castellelex Ltd is a British manufacturer specialising in NBC/CBRN air filtration and Collective Protection systems for private, commercial, government and specialist applications.

Castellelex systems are manufactured in the United Kingdom and supplied internationally for applications ranging from private family fallout shelters and bunkers to larger commercial and civil-protection projects.



ICI
International
CBRNE
INSTITUTE

**A common roof
for International
CBRNE
First Responders**



Rue de la Vacherie, 78
B5060 SAMBREVILLE
(Auvelais)
BELGIUM

info@ici-belgium.be | www.ici-belgium.be