

WRITTEN TESTIMONY OF KURT GAUDETTE

SENIOR VICE PRESIDENT, DRAGOS, INC.

Before the Subcommittee on Disaster Management, District of Columbia, and Census

Committee on Homeland Security and Governmental Affairs

United States Senate

“Rogue AI: Securing the Homeland Against AI Agent Attacks”

September 30, 2026

Chairman Hawley, Ranking Member Kim and distinguished members of the Subcommittee, thank you for providing me the opportunity to testify before you today. My name is Kurt Gaudette and I am the Senior Vice President of Intelligence & Services at Dragos, Inc. a company that is singularly focused on protecting industrial control systems (ICS) and operational technology (OT). I lead a team that hunts, discovers, and tracks threats to ICS and OT networks. I served over 30 years in the U.S. military and the intelligence community as a senior officer in the U.S. Air Force and as part of the Senior Executive Service at the National Security Agency, where I built a global enterprise program focused on weapons, space and disruptive technologies.

Over the past decade, Dragos has built one of the largest available bodies of real-world OT data drawn from our threat intelligence, incident response, and vulnerability research work inside these environments. It gives us a distinct, evidence-based perspective on how AI agents are already affecting the threat landscape facing the nation’s critical infrastructure, and also what they might be able to do in the future.

Today, our nation’s critical infrastructure owners and operators find themselves on the front lines. They must defend their networks from a growing number of cyber threats from both state actor and criminal groups in a rapidly evolving threat environment, including increasing AI-assisted attacks. To successfully defend these networks requires both an understanding of the environment and investment in the right resources. My testimony focuses on three key points that are relevant to the Subcommittee and this hearing’s focus.

- First, the cyber threat landscape for operational technology and industrial control systems is evolving more quickly than ever before. For more than four decades adversaries in the OT space were largely state-sponsored actors conducting reconnaissance and gathering information. There were only a handful of high-cost, high-consequence attacks on OT environments. More recently, digitalization, connectivity and uniformity in OT is adding risk. Adversaries can now develop scalable and reusable capabilities to target OT. State-sponsored actors have shared capabilities with non-state actors such as hacktivists and criminals, quickly increasing the number of threat groups capable of targeting ICS and OT networks. And this year, for the first time, Dragos observed AI-assisted attacks specifically targeting OT infrastructure.
- Second, AI is changing the OT threat landscape. It lowers financial and knowledge barriers for adversaries. Work that used to require a skilled, patient operator with

specialized industrial knowledge can increasingly be attempted by a much larger population of less-skilled actors, in a fraction of the time, using commercial AI models. These same models enable vulnerability discovery at scale. Organizations need to be prepared for a massive influx of vulnerabilities very quickly and ready to ruthlessly prioritize. Within the context of the evolving threat landscape and proliferation of OT-focused threat groups and vulnerabilities, prevention-only strategies are not sufficient. Defenders must be able to detect and respond to anomalies as well.

- Finally, models will keep improving, but even in an environment where AI can extend the scale and reach of adversaries, AI-assisted attacks do not fundamentally change how defenders actually protect their networks. We have the basic security playbook. In every case Dragos has directly investigated, the underlying adversarial techniques are unchanged. What has changed is the urgency with which organizations need to act. Those without even the basic controls in place are already at a much greater risk and that risk is increasing rapidly.

I. Evolution of the Operational Technology Threat Landscape

OT systems are the specialized computers and networks that interact with the physical environment to do things like control the pumps or chemical levels at a water treatment facility, or a circuit breaker at an electrical substation. The distinct purpose of these systems informs how risks and threats to the system are defined and managed. The impact of a breach or compromise is different as well, especially across information technology (IT) and OT environments.¹ For IT systems, cybersecurity largely focuses on system and data security. In OT environments, if adversaries gain access, they can potentially disrupt critical operations, produce physical damage, or even cause safety-related incidents that lead to equipment damage or loss of life. Defenders must maintain availability while also considering security. Many OT critical components can't be turned off and downtime cannot be planned regularly for things like patching and updates. All of this has shaped the approach to OT cybersecurity.

Over the last fifty years, we can observe three distinct eras of OT cybersecurity defined by the types of attacks and adversaries observed in the real world. First, from the 1980s through as late as 2021, adversaries in operational environments were overwhelmingly intelligence agencies. Throughout this time, adversaries were observed mostly attempting to access OT environments and industrial control systems networks to learn about the systems, determine what value there might be in reaching them and simply observing. Destructive or offensive operations were rare because they were difficult and resource intensive. The environments themselves were hard targets. ICT and OT were largely disconnected from other networks. The infrastructure was also complex and heterogenous with little in common between two facilities even in the same industry, making it difficult for adversaries to create attacks that caused disruption or physical destruction. A refinery in one country had almost nothing in common with an electric transmission substation or a water treatment plant somewhere else, so every attack with real

¹ <https://www.sans.org/white-papers/36297/>

physical consequence was expensive, specific, and slow to build. Scaling an attack was even harder. The result was a long era of low-frequency, high-consequence activity.

All of this shifted as OT systems became more homogeneous, digital, and connected. Now, under the hood of controllers from different vendors sit common software stacks, common protocols, and cookie-cutter installations at facilities across geographic locations and industries. Threat groups can develop capabilities that target devices commonly used in OT environments across sectors and have found new ways to access and manipulate them to cause disruption and pose safety risks. Adversaries also can model, train, and scale attacks in ways they were not able to in heterogeneous environments. In early 2022, Dragos analyzed and named PIPEDREAM, the first cross-industry, reusable and modular framework capable of causing physical effects across multiple sectors. The PIPEDREAM toolkit had the capability to impact devices that control critical infrastructure in different sectors, including devices that manage electrical systems, oil and gas pipelines, water systems, manufacturing plants, and even the control systems in military assets such as unmanned aerial vehicles and naval ships. PIPEDREAM also takes advantage of native functionality in the software and network protocols available cross-industry, so it cannot be patched away.²

By 2024, Dragos could assess with high confidence that state actors were arming non-state actors, including hacktivists and criminal groups, with OT-specific knowledge and the infrastructure to manipulate things such as controller logic and physical processes.³ Almost overnight, non-state actors, who had previously only been observed with capabilities such as defacing internet-facing devices, had program logic controller (PLC) level implants and tools for ladder logic manipulation. This development rapidly expanded the risk picture for critical infrastructure, increasing the number of threat groups with the capabilities to specifically target OT and making it even more imperative for organizations to invest in and implement OT security controls. In 2026, AI has changed the OT threat landscape again.

II. How Does Artificial Intelligence Change for Operational Technology Cybersecurity?

AI is not a novel topic and conversations about its use cases and risks, particularly in the hands of our adversaries, have been taking place for years. However, only recently have we been able gain a field view of what AI actually changes for OT security, including where the threat is real and what defenders need to know. Over the last nine months, Dragos has worked first-hand with frontier AI models, watched its own red team and offensive specialists put those models to work, hardened its own products against them, and investigated incidents where adversaries were already using AI in the early stages of their operations. Our observations fell into three distinct categories.

AI Escalates IT-Focused Actors Toward Operational Technology Targets and Lowers the Barrier to Entry at the Network Edge

² https://hub.dragos.com/hubfs/116-Whitepapers/Dragos_ChernoviteWP_v2b.pdf

³ <https://hub.dragos.com/ot-cybersecurity-year-in-review-2025>, p. 38

First, AI lowers both knowledge and financial barriers for adversaries to attack critical infrastructure, but it has not, to date, raised the ceiling of what those methods can accomplish against a well-defended OT environment. Work that used to require a skilled, patient operator with specialized industrial knowledge can increasingly be attempted by a much larger population of less-skilled actors, in a fraction of the time. When compared to the number of threat groups targeting IT infrastructure, the number of groups who have the intent, knowledge and resources to attack OT infrastructure is much smaller. Dragos tracks at least 26 such groups.⁴ The impact could be greater, but this barrier has likely insulated our communities from a greater number of successful attacks with consequences. However, we now have examples of models pointing IT-focused adversaries toward OT networks.

In early 2026, researchers identified a large-scale breach of multiple Mexican government IT networks.⁵ Dragos independently reviewed operator logs and technical artifacts recovered from that campaign and found that the attacker had used Anthropic's Claude AI model to support technical execution of the intrusion, including reconnaissance, tool development, and lateral movement. Dragos's specific focus was an intrusion into the enterprise IT environment of a municipal water and drainage utility, Servicios de Agua y Drenaje de Monterrey (SADM).

Dragos found that once inside SADM's IT network, the AI model identified an internally accessible interface to the utility's operational technology, without being specifically directed to look for control-system infrastructure. It assessed that interface as a high-value target because of its proximity to critical infrastructure, and generated and executed a tailored credential attack intended to breach the boundary between the utility's IT and OT environments. That attack failed. Roughly 2,800 login attempts were rejected, and the intrusion never obtained access to, or visibility into, any underlying industrial system. The IT/OT boundary held.⁶

This case matters less for what the attacker achieved than for what the AI model volunteered. An attacker with no apparent OT-specific skill or intent was, in effect, steered toward a critical-infrastructure target by the tool itself. Had default vendor credentials or weak passwords been in place at that interface, Dragos assesses the intrusion could plausibly have progressed further into the control-system environment.

Dragos has also observed AI-assisted intrusion activity against internet-facing firewalls and virtual private network (VPN) appliances that connect directly to operational networks. The infrastructure need not be OT-specific to create OT risk. In one such case affecting a gas utility, an attacker used a commercial AI coding assistant as an interactive intrusion tool to access a compromised firewall, gaining entry with valid administrative credentials rather than by exploiting a vulnerability. During that operation, the AI assistant took an incorrect configuration action that rendered the device unresponsive, disrupting the utility's own site-to-site connectivity. Dragos found no evidence this activity reached or affected any ICS/OT asset, and no evidence the operator had any ICS-specific skill or objective. The attacker never gained routed access to the utility's internal network, and once the AI's configuration change took the firewall offline, the operator did not regain access. More broadly, the case shows that adversary capability

⁴ <https://www.dragos.com/ot-cybersecurity-year-in-review#download-report-2026>

⁵ <https://www.latimes.com/business/story/2026-02-26/hacker-used-anthropics-claude-ai-to-steal-mexican-government-data>

⁶ <https://www.dragos.com/blog/ai-assisted-ics-attack-water-utility>

increasingly need not be ICS-specific to create operational risk. What is key is operational influence: if a system impacts the control loop (or physical processes), it is part of the operational environment, whether or not it was built for OT.

Dragos is not aware of a confirmed case of AI-generated malware causing physical impact to an industrial process. Claimed AI-generated ICS malware that has circulated publicly has, on inspection, been non-functional. AI models pointed at vulnerability discovery in real ICS/OT software have produced a meaningful number of false positives alongside genuine findings, and independent research suggests AI-assessed severity runs high more often than not. In both intrusions described above, the AI model introduced no new tradecraft. Neither reached OT. In the water utility case, the credential attack failed because none of the passwords it tried were valid. In the gas utility case, the attacker controlled the perimeter firewall but never reached the internal network behind it.

AI Accelerates Vulnerability Discovery and Exploit Development

Published research also demonstrates how AI models are compressing the time and expertise historically required to turn a known but hard-to-exploit vulnerability into a working attack against real industrial hardware. In September 2026, researchers at a security vendor used Anthropic's Claude, first Sonnet 4.6, then Opus 4.6, to port a previously developed remote-code-execution exploit from one WAGO programmable logic controller model to a related but distinct model still in active industrial use. After working through several days of dead ends, the model produced two independent working payloads within twelve minutes once it identified the correct approach, at a cost of a few hundred dollars in compute.⁷

The underlying flaw is a pre-authentication buffer overflow first disclosed in 2021, for which no firmware patch exists on the affected device family. Dragos assessed this vulnerability class at the time of its original disclosure as difficult to exploit against a specific device without significant bespoke research, but this experiment shows that assumption eroding. The security vendor that conducted the experiment was clear that a skilled human researcher could still have ported this particular exploit faster, and without the one accidental step that permanently destroyed the test device during a follow-on session. The more durable takeaway, in Dragos's assessment, is the trend. The population of actors who can adapt a known vulnerability across a family of fielded, unpatchable industrial devices is growing, while the specialized research hours that once made that work a meaningful barrier to entry are shrinking.

Models enable vulnerability discovery at scale, including in industrial environments. Pointed at industrial products inside Dragos control system ranges, with the right harnesses and tooling, these models found combinations of vulnerabilities that are not easy for human analysts to surface. Not a handful. Dozens of strong zero-days across major original equipment manufacturers (OEMs), with hundreds into the thousands of additional vulnerabilities. This is important because the OT defender community is about to be overwhelmed by sheer volume of vulnerabilities. In environments when patching is selective and complicated, even if the community were able to rapidly assign each of these vulnerabilities a CVSS score, it would not be enough to help defenders prioritize. Instead, organizations need to know what vulnerabilities actually matter. Dragos applies a “Now, Next, Never” model for exactly this reason. Roughly 20

⁷ <https://www.forescout.com/blog/can-ai-create-plc-attacks-yes-but-it%E2%80%99s-not-that-easy-yet/>

to 30 percent of vulnerabilities add no meaningful operational risk and belong in the Never bucket. In the context of industrial and operational environments, 60 to 70 percent require compensating controls or a place on the list under Next. Only about three to six percent of vulnerabilities demand immediate action because they enable safety or control manipulation or are actively being exploited. As AI multiplies the vulnerability count exponentially across both legacy and modern equipment, prioritization becomes the difference between a manageable security program and an impossible one.

Additionally, the same discovery capability is a tool for defenders willing to use it. Vendors and OEMs who invest the money to run these models against their own products before shipping will find their flaws before adversaries do. Many will not make that investment, and a great deal of under-tested equipment will continue to enter operational environments. The vulnerabilities that matter most are not usually the ones deep in a Programmable Logic Controller (PLC). They are the ones closer to the perimeter of the automation environment, in the jump hosts, firewalls, and access points, where a flaw that grants access gets abused quickly. Dragos tracks a threat group that, before adopting AI, could reverse-engineer a freshly released vulnerability and build a working exploit within 24 to 48 hours. In an AI-assisted world, that window shrinks toward minutes. A day or two was already more than most organizations could keep pace with.

Prevention-Only is not a Viable Strategy

The third pattern we observed is that prevention-only cybersecurity strategies are not enough. Many infrastructure owners and operators have adopted a prevention-only posture because most of the guidance and recommendations focus on prevention. Dragos analysis of dozens of standards found that well north of 80 percent, in some cases as high as 98 percent, of the guidance is preventative in nature: antivirus, patching, passwords, access controls, firewalls, segmentation. Prevention remains necessary, but is not sufficient. OT environments have become more homogeneous, more connected, and more digital. At the same time, the number of adversaries targeting OT is growing and commercial AI models are both pointing adversaries to OT and supplying them with specific information for how to breach the IT-OT boundary. A control set that stops at prevention will not be enough. It always fails eventually, and when it does, organizations without detection, monitoring, and response capabilities may not be able to detect a breach quickly or be able to determine the causes.

It is important to note, however, that not every adversary becomes dangerous to OT environments because AI exists. Most threat groups target IT systems and many use unsophisticated attack techniques aimed at targets of opportunity. Dragos tracks at least 26 groups that target OT specifically, and within that set, only four or five are of greatest concern. These are the teams that have run OT operations against our critical infrastructure for fifteen to twenty years and have almost certainly retained data from those operations. That accumulated operational data (training data), layered on top of a frontier model, is what turns a capable adversary into one with reach and scale the OT community has not previously faced. Frontier models are strong starting points for adversaries, but they need specific context to matter for OT environments. A handful of adversaries hold a comparable data advantage, and their ability to apply it will give them extraordinary reach. Defenders need to be prepared.

III. Critical Infrastructure Defense in a Rapidly Changing Threat Landscape

Models will keep improving, and OT context will become more widely available as research labs and universities publish on how controllers can be attacked. This is training data. In the future, it is not unreasonable to expect models with enough context to operate in automation environments with less human interaction. When Dragos' own red team applied its knowledge to a model, that model moved through the perimeter, mapped the automation environment passively and accurately, identified the exact original equipment manufacturer and system, and found the right exploits to reach the controllers. That capability is not in public models today, but it likely will be soon.

This only reinforces that prevention alone is not enough. But even in an environment where AI can extend the scale and reach of adversaries, we have the basic security playbook. AI-assisted attacks do not fundamentally change how defenders actually protect their networks. OT compromise still overwhelmingly comes down to known tactics, techniques, and procedures for misoperating equipment and using operational systems against themselves. Models still need the training data. What AI does change is the urgency with which organizations need to act. Those without even the basic controls in place are already at a much greater risk and that risk is increasing rapidly.

Organizations should ensure they have implemented the foundational OT security controls, outlined in the SANS Five ICS Cybersecurity Critical Controls: 1) ICS-Specific Incident Response Plan 2) Defensible Architecture 3) ICS Network Visibility and Monitoring 4) Secure Remote Access 5) Risk-Based Vulnerability Management⁸

As the frequency of attacks increases, especially AI-assisted attacks, it becomes more important to quickly detect anomalies and understanding what is happening in your OT systems. Today, Dragos estimates fewer than 10 percent of OT environments worldwide have visibility and monitoring in place.⁹ This also means they cannot determine root cause when something goes wrong, whether the cause was a cyberattack, a malfunction, or human error. It's a blind spot, and AI only raises its cost, because faster incidents are harder to reconstruct after the fact when the underlying visibility was never there.

This is not a minor technical gap. Without root cause analysis, incident responders cannot confidently restore operations, cannot warn other operators of an ongoing threat, and cannot learn from what went wrong to prevent a repeat. It also has a national security dimension. An adversary who can operate inside a critical infrastructure network without being detected gains what amounts to plausible deniability, which complicates any government's ability to respond decisively.¹⁰ AI makes this gap more consequential.

Additionally, organizations and agencies should plan for failure scenarios by understanding intelligence-driven threat scenarios and consequence-scenarios that are already occurring. You cannot protect yourself against the unknowns if you cannot handle the knowns. When used

⁸ <https://www.sans.org/white-papers/five-ics-cybersecurity-critical-controls>

⁹ <https://www.dragos.com/ot-cybersecurity-year-in-review#download-report-2026>

¹⁰ <https://www.weforum.org/stories/cybersecurity/dangerous-blindspot-in-infrastructure-cybersecurity/>

responsibly and with safeguards in place, defenders can scale their efforts with AI as well. For defenders, models need the field knowledge, the assessment history, and the incident response experience that only comes from years of operational work. A great model that uses bad data is not likely to produce a usable outcome. Good data, good analysts, and a capable model together can produce results that surpass human-only efforts. Assessments that took two to three weeks can be compressed to two to three days. Threat hunts that took twenty hours can be done in minutes. The model does not replace the analyst. It elevates the analyst when bounded carefully to guard against hallucination. Organizations should establish governance processes for AI adoption that balance innovation with security and safety and apply the same testing, validation, and continuous assessment rigor to AI deployment that they have historically applied to control systems. Leaders also need to understand the significant risk to the business in letting competitive or economic pressures override the engineering discipline that has kept operations safe and reliable

IV. Recommendations

The continuously evolving OT threat landscape, including the growing use of AI agents by threat groups to scale and scope attacks, presents additional risk to critical infrastructure. There are practical measures that should be considered to help infrastructure owners and operators reduce this risk and strengthen overall resilience. The following recommendations are intended to address key areas of vulnerability and provide actionable steps for supporting effective risk management

Information Sharing Authorities

Information-sharing and convening authorities, as well as liability protections that encourage private-sector threat sharing, remain essential to the kind of cross-sector communications that let agencies, operators, service providers and others rapidly share critical warnings and indicators. This becomes even more important in a world with growing AI-assisted activity, such as the cases described above. We must move at the speed of adversaries and technology. Continued support and reauthorization of these programs in some form, including CISA 2015, provides operators and agencies with a necessary tool. Lapses would slow the flow of information between operators, vendors, and the government at a time when we cannot afford it.

Resourcing for Small Utilities

Programs that put resources directly in the hands of under-resourced operators deliver a direct return against the asymmetric threat of AI-assisted cyber attacks. This includes both industry support and grant funding to reduce technology debt and enable foundational OT cybersecurity controls such as network segmentation and network visibility. Grant programs that can put resources directly in the hands of operators, including the State and Local Cybersecurity Grant Program and the Rural and Municipal Utility Advanced Cybersecurity Grant and Technical Assistance Program can be sustained and expanded where possible to combat the security gap these organizations face. As AI compresses attacker timelines, the value of closing this resourcing gap only increases.

A Unified National Operational Technology and Industrial Control System Incident Response Plan

Asset owners and operators need a clear, consistent way to reach federal responders before, during and after an incident that involves OT. They also need to know what support they can expect to receive and find consistent expertise and credentialing among responders from state to state. A unified national incident response plan, specific to ICS and OT, would address each of these concerns. The technical plan for this is not difficult. One has already been developed and piloted at the 91st Cyber Brigade out of the Virginia Army National Guard with a participating utility. It demonstrated that capability can be quickly scaled prior to incident and delivered to even small utilities. Now that there is proof of concept, this plan could be expanded nationally, with CISA in the coordinator role to work with sector risk management agencies, the Federal Bureau of Investigation and other relevant federal and state agencies. Currently, however, there is a patchwork of federal authorities and budget lines that do not align to how incident response occurs during a real-world incident. They were developed independently over the years to address capability needs, without the benefit of considering how operational integration looks during an incident. The Congress can address this by requiring a coordinated, national plan and aligning funding and authorities across agencies as outlined in the plan.

Conclusion

In conclusion, the cyber threat landscape for operational technology and industrial control systems is evolving more quickly than ever before, especially given the demonstrated capability for commercial AI models to assist in cyberattacks on OT environments. We are no longer facing the threat of low-frequency, high-consequence attacks on a handful of targets, but an era when all critical infrastructure are targets. AI is already shifting the operational technology threat environment. It lowers both knowledge and financial barriers for adversaries and has directed IT-focused threat groups to OT networks. On the defender side, as models enable vulnerability discovery at scale, organizations will need to quickly and reliably determine which matter most to prioritize remediation.

The number of adversaries targeting OT is growing and commercial AI models are pointing adversaries at OT. At the same time, OT environments have become more homogeneous, more connected, and more digital. Prevention-only strategies are no longer sufficient. Organizations need to prepare for breach and be able to detect and respond; however, defense is doable. Even as AI can extend the scale and reach of adversaries, we have the security playbook and AI-assisted attacks do not fundamentally change how defenders actually protect their networks. Our challenge across government and industry is to ensure that all of our critical infrastructure, from the smallest water utility to the largest critical manufacturer, is able to implement and invest in foundational security controls.

I sincerely thank the subcommittee for providing me the opportunity to testify today and look forward to your questions.