INFINITUM CORPORATION

Dead Reckoning

AI for Security Operations in a Jammed, Spoofed, Watched World

Practical, Integrated Steps for Mid-Size Organizations

TRUST NO SINGLE SIGNAL.

William Aaron Turner, Founder & President
Infinitum Corporation, Washington, DC
t.me/infinitumcorporation
www.theinfinitumcorporation.com
info@theinfinitumcorporation.com

Whitepaper 01

OCTOBER 2026

Dead Reckoning: AI for Security Operations in a Jammed, Spoofed, Watched World Practical, Integrated Steps for Mid-Size Organizations

Infinitum Corporation Whitepaper No. 1 · October 2026 · **Released October 25, 2026**

Neutrality statement. This paper draws general lessons about technology and conflict from public sources. It takes no side and makes no claim about any party.

© 2026 Infinitum Corporation. All rights reserved. No part of this publication may be reproduced or distributed without written permission, except for brief quotations with attribution.

DISCLAIMER

This whitepaper is provided for general information only. It is not legal, regulatory or professional advice for any specific situation, and it does not create a client relationship. Statistics are taken from the public sources cited and remain the property of their publishers; they describe the samples studied by those publishers and are not benchmarks or predictions for any particular organization. Infinitum Corporation makes no promise or guarantee about the security outcome any organization will achieve.

This paper is strictly analytical and defensive. References to irregular warfare, gray-zone competition and hybrid threats are used as an analytic lens for defense and resilience. Nothing in this paper is instruction for offensive activity. Any security testing referred to here means authorized, scoped testing performed with the written permission of the system owner.

Scenarios in Part IX are illustrative and hypothetical. They do not describe any real organization, client or incident.

Trademarks and framework names belong to their respective owners. Citation of a source does not imply that its publisher endorses this paper or Infinitum Corporation.

CONTACT

Infinitum Corporation, Washington, DC · 1629 K Street NW, Suite 300
(202) 980-9300 · https://www.theinfinitumcorporation.com
info@theinfinitumcorporation.com
Telegram: https://t.me/infinitumcorporation · X: https://x.com/infinitumceo

All data comes from the public record. All conclusions are mine. — W.A.T.

Contents

Executive summary 4

The problem 7

The modern battlefield 10

Analysis 16

Part I: The irregular warfare lens 16
Part II: Adversary adaptation 23
Part III: The evidence on AI for defenders 26
Part IV: The use cases, one by one 28
Part V: Governance and AI risk 34
Part VI: Securing the AI itself 36
Part VII: People and the workforce 39
Part VIII: Metrics that matter 42
Part IX: Illustrative scenarios 44

Recommendations 46

Conclusion 51

About Infinitum Corporation 53

References 54

Glossary 56

Executive summary

By William Aaron Turner, Founder and President, Infinitum Corporation.

The signal is gone.

Your navigation says one thing. Your clock says another. The voice on the phone sounds exactly like your chief executive.

That is not a scene from the front line. It is a description of the world your security team works in, and the public record of the modern battlefield shows where that world is heading.

This is not a hype piece about AI. It is a practical look at where AI helps a security team, where it hurts, and how a mid-size organization can start without losing control.

I read security problems through one lens: irregular warfare. That is the study of how states, proxies and other groups compete below the threshold of open war, using indirect, deniable and asymmetric means. You don't have to care about military doctrine to benefit from it. You only have to notice that the people attacking your network are behaving the way that doctrine describes.

They probe. They hide in normal activity. They pick the cheapest way in. They wait.

Then look at what public analysts say about the modern battlefield. Constant observation. Cheap mass against expensive systems. Signals jammed and spoofed. A cycle of measure and countermeasure that turns in weeks (CSIS, 2025a; CSIS, 2025b). My shorthand: the grind of 1916 fused with science fiction.

That is a comparison, not a claim that your network is a front line. Nobody is shooting at your servers. This paper draws general lessons about technology and conflict from public sources. It takes no side and makes no claim about any party. The lessons transfer, and I spend a chapter on seven of them.

The title is a navigation term. *Dead reckoning* is how you find your position when the signal is jammed or spoofed: last known position, heading, speed, time, and a second source to check against. It is the posture I want your security operations to take.

And now they have AI.

Here is the short version. Three findings from recent research stand out.

First, speed. Verizon's 2026 Data Breach Investigations Report (DBIR) found that exploiting software vulnerabilities is now the most common way into a breach, while only 26% of the most critical known-exploited vulnerabilities were fully fixed (Verizon, 2026).

Second, AI works for defenders too. Organizations that used security AI and automation extensively averaged lower breach costs and faster response in IBM's 2026 study (IBM and Ponemon Institute, 2026). Key word: averaged. And most organizations have not adopted these tools widely yet.

Third, AI is also a new attack surface. The research points to plain gaps in access control and governance, not exotic technical failures.

Put those three together and you get the core problem of this paper. Your adversary is getting faster and cheaper. Your team is not getting bigger. AI can close part of that gap. Adopted carelessly, it opens a new one.

The irregular warfare lens adds something the statistics alone do not. It explains why the problem looks the way it does.

Gray-zone actors stay below thresholds. In your environment, that means activity designed to look like normal administration. Hybrid campaigns blur attribution. In your environment, that means you will often not know who is behind an intrusion, and you will have to act anyway. Irregular actors impose costs asymmetrically. In your environment, that means an attacker spends little to make you spend a lot. And the doctrine on countering hybrid threats keeps returning to one answer: resilience. Institutions that can absorb a blow, keep working and recover are much harder to coerce.

That is the standard I hold an AI program to. Does it make your organization more resilient, or does it simply make it busier?

My recommendation is simple and not glamorous: govern first, measure a baseline, pilot with analysts in the loop, point AI at exposure, secure the AI itself, and widen automation only as the evidence earns it. Technology, process and people, working as one system. That is what practical, integrated security looks like.

This paper is organized to support that recommendation.

- **The problem** sets out the pressures: attacker speed, slow remediation, expensive breaches and thin teams.
- **The modern battlefield** draws seven lessons for defenders from public analysis of the modern battlefield, from CSIS and the bodies tracking satellite-navigation interference.
- **Part I** introduces the irregular warfare lens, with definitions taken from U.S. joint doctrine, CSIS, RAND, NATO and the European Union.
- **Part II** looks at how adversaries adapt, using a public U.S. government advisory as a worked example.
- **Part III** reviews the evidence that AI helps defenders, and where adoption is uneven.
- **Part IV** analyzes six security operations use cases in depth: alert triage, threat-intelligence enrichment, vulnerability prioritization, log analysis, reporting, and deception and early warning.
- **Parts V and VI** cover governance under the NIST AI Risk Management Framework and the job of securing the AI itself.
- **Parts VII and VIII** address people and metrics.
- **Part IX** walks through three illustrative scenarios. They are hypothetical, and labeled that way.

- **Recommendations** give a phased roadmap for the first 90 days and the first 12 months, plus a maturity self-assessment you can run with your team.

A word on sources. Every statistic carries its source. Every definition of doctrine is quoted or paraphrased from a public document I opened and checked. Where I offer judgment, I say so.

All data comes from the public record. All conclusions are mine.

The problem

Three pressures are landing on mid-size organizations at once: faster attackers, slower fixes and thin teams. It is the arithmetic of attrition. The side that can sustain the grind holds the ground.

Let's take them one at a time.

Attackers are walking in through unfixed software

Verizon found that vulnerability exploitation was the initial access vector in 31% of breaches, up from 20% in the prior year's report. It overtook credential abuse (13%) as the leading vector (Verizon, 2026).

Verizon also reports that AI is helping threat actors speed up the time it takes to exploit known vulnerabilities (Verizon, 2026).

Think about what that means in practice.

A vulnerability is published. A patch is released. The clock starts. On one side, an attacker who can now move from disclosure to working exploitation faster than before. On the other side, your change-management calendar.

The race is not fair. It was never meant to be.

Fixes are not keeping pace

Only 26% of critical vulnerabilities listed in CISA's Known Exploited Vulnerabilities (KEV) catalog were fully remediated in 2025, down from 38% the year before. The median time to full remediation rose to 43 days from 32 days. The median organization had 16 KEV vulnerabilities to deal with, up from 11 (Verizon, 2026).

Let that sink in.

These are not theoretical weaknesses. KEV is CISA's list of vulnerabilities known to have been exploited in the wild. CISA advises organizations to use the catalog as an input to their vulnerability prioritization framework (CISA, KEV catalog). Yet most of the critical ones were not fully fixed, and the time to fix them got longer.

The direction is wrong on every measure in that paragraph. Fewer fixed. Slower to fix. More to fix.

Breaches are expensive, and slow to find

IBM's 2026 Cost of a Data Breach Report, conducted by Ponemon Institute, put the global average breach cost at USD 4.99 million, a 12% increase and a record. The mean time to identify and contain a breach was 247 days (IBM and Ponemon Institute, 2026).

Two hundred forty-seven days.

That is roughly eight months in which an intruder may be learning your environment, your people and your habits before you have both found and contained them.

The AI dimension is already showing up in the numbers. More than one in four organizations hit by a malicious attack said it was AI-driven, a 56% increase over the previous year, and AI-driven attacks added an average of USD 1 million to the cost of a malicious breach (IBM and Ponemon Institute, 2026).

And the people are stretched

The 2024 ISC2 Cybersecurity Workforce Study surveyed 15,852 practitioners and decision-makers. 67% reported a staffing shortage. 90% reported one or more skills gaps. Among organizations with skills gaps, 22% reported a material breach, compared with 17% of all respondents who shared that information (ISC2, 2024).

That last comparison matters. Skills gaps are not just an HR problem. In the ISC2 data, they travel with worse security outcomes.

A caution on the numbers

Now, I know what you're thinking: these studies are global, not mid-size.

You're right.

None of them is limited to mid-size organizations, and they use different samples and different years. Treat the numbers as direction, not as a benchmark for your company. IBM itself notes that its sample is not statistical and leans toward organizations with more mature security programs.

I will keep repeating this point through the paper, because it is the most common way security statistics get misused. A survey average tells you which way the wind is blowing. It does not tell you how strong the wind is at your address.

Why this is a capacity problem

For a mid-size organization, all of this adds up to a capacity problem.

More vulnerabilities to fix. More alerts to read. More AI-enabled threats to recognize. The same small team.

Mid-size organizations sit in an awkward place. They are big enough to be worth attacking, with real data, real money and real connections to larger partners. They are rarely big enough to staff a full security operations center around the clock, run a dedicated threat-intelligence team, and maintain a separate vulnerability management function.

So the same few people triage alerts in the morning, chase patches in the afternoon and write the board report at night.

That is the opening for AI. It can take some of the reading, sorting, summarizing and drafting off those people.

It is also why adopting it carelessly can make things worse. A tool that produces confident, wrong answers at machine speed does not reduce workload. It moves the workload to cleanup, and it moves the risk to the places nobody is checking.

The question this paper answers

So the question is not whether to use AI in security operations. Attackers have answered that question for us.

The question is how a mid-size organization adopts it in a way that makes the whole program stronger: more resilient, better prioritized, harder to surprise.

To answer that well, it helps to understand the kind of contest you are in. That is where irregular warfare comes in.

The modern battlefield

Think of the grind of 1916 fused with technology that looks like science fiction.

Persistent observation. Drones overhead. Satellites watching. Radios jammed. Positions spoofed. A cycle of measure and countermeasure that turns in weeks.

That is my shorthand for the modern battlefield. I use it as a comparison and nothing more. I am not a war correspondent, and this paper is not a military study. I am a security consultant reading public analysis, and asking one question: what does this teach the people who defend organizations?

Here are the ground rules.

This paper draws general lessons about technology and conflict from public sources. It takes no side and makes no claim about any party. I do not assess any state, army or leader. I leave out casualty, loss and performance figures for every side, because they are contested, because they are estimates, and because they are not what a security team needs. I keep the lessons at the level of technology and phenomena: drone saturation, electronic jamming and spoofing, sensor transparency, cheap mass, logistics under observation and decentralized decisions.

The main sources are the CSIS report *War and the Modern Battlefield* (2025) and European and aviation authorities that track interference with satellite navigation. I opened each document I cite. Where a source gives an opinion or a forecast, I say so. And I stay strategic. Nothing here is a tactic, a targeting method or a how-to.

The title of this paper comes from navigation. *Dead reckoning* is what you do when the signal is gone. You know where you were. You know your heading, your speed and how much time has passed. You work out where you are, and you check it against anything else you can trust.

That is the posture I want you to take into security operations. Because the signals you depend on are now being jammed, spoofed and watched.

Seven lessons

Lesson 1: Persistence beats brilliance

What public analysis shows. CSIS observes that persistent coverage from space assets is making the battlefield more transparent and diminishing strategic surprise (CSIS, 2025b). It adds that this transparency does not by itself guarantee restraint or deterrence (CSIS, 2025b).

That is CSIS. Here is my reading of it. When surprise is hard to achieve, contests drag. The side that can sustain effort over time matters more than the side with one clever move.

What it means for you. Your contest is also a long one. IBM and Ponemon Institute put the mean time to identify and contain a breach at 247 days (IBM and Ponemon Institute, 2026). ISC2 found that 67% of respondents reported a staffing shortage and that budget was the top cause of skills gaps (ISC2, 2024). Attrition hits your side too: tired analysts, thinner budgets, a backlog that does not close.

So plan for a campaign, not an incident. Fund the second year, not just the first quarter. Protect your people from burnout the way you protect systems from failure. A team that is exhausted by month six is a position that will not hold.

Where AI fits. AI is a way to spend fewer analyst hours on the repetitive work that wears people down. It does not remove the need to staff the line.

Lesson 2: Cheap mass versus exquisite systems

What public analysis shows. CSIS puts the economics in one line: "If a $500 drone can destroy a multimillion-dollar tank, so, too, can a USV swarm destroy a fleet of ships" (CSIS, 2025a). That is an illustration of an asymmetry in cost, not a measurement of how often it happens.

CSIS is also clear that no tool is a magic bullet. In its space chapter it says electronic warfare is not a magic bullet against drones or munitions, because technologies and tactics continue to evolve to mitigate its effects (CSIS, 2025b). Cheap mass invites cheap countermeasures, and the contest goes on.

What it means for you. The adversary's cheap mass is phishing kits, automated scanning and commodity ransomware services. ENISA describes Phishing-as-a-Service as automation that "paves the way for attackers regardless of their experience" (ENISA, 2025). IBM puts the imbalance in a sentence: attacks "can be launched for thousands while breaches cost millions" (IBM and Ponemon Institute, 2026).

Most of those attempts fail. A few land. That is the model. You will not stop every attempt. You make the average attempt fail, and you make the successful one small.

Do not answer cheap mass with one exquisite tool. Answer it with layers, and with cheap, repeatable, automated controls that run all day: patching known-exploited vulnerabilities, multifactor authentication, email filtering, segmentation.

Where AI fits. This is where AI belongs: cheap mass on your side. Triage, summarization and enrichment at volume, with a person on the decision.

Lesson 3: You are being watched

What public analysis shows. CSIS says forces "must prepare to fight in full view of the enemy," and points to cheap commercial overhead imagery and crowd-sourced intelligence as part of the picture (CSIS, 2025a). In its space chapter, CSIS says military forces, especially those at fixed sites, "must assume that they will be seen and their movements and emissions detected," which calls for renewed emphasis on operational security, deception and spectrum discipline (CSIS, 2025b).

Observation changes behavior. When everything visible can be seen, you decide what is worth showing.

What it means for you. Your organization is observed too. MITRE ATT&CK lists Reconnaissance as the first tactic in its enterprise matrix, with techniques such as active scanning and gathering information on victim identity and organization (MITRE ATT&CK). That means adversaries study your internet-facing systems, your staff, your suppliers and your business rhythm before they touch anything.

Work out what is visible. Which systems answer on the internet? Which staff names, roles and email formats are public? Which vendors are named on your website? Then reduce what does not need to be there. Concealment has a place in cyber defense: fewer exposed services, fewer stale accounts, fewer forgotten servers.

Then build your own sensors. Logs are your eyes. If you cannot see your own network, you are the one fighting blind.

Where AI fits. AI can help you read the sensors you already have, such as logs, alerts and external exposure data. It cannot help you see what you never collected.

Lesson 4: Signals can be jammed and spoofed

What public analysis shows. CSIS calls electronic warfare one of the most dominant features of the modern battlefield. It describes these systems as a "temporary and reversible" way to attack satellite navigation, communications and surveillance collection, and notes that jamming and spoofing of GPS signals has been detected in several regions around the globe (CSIS, 2025b). CSIS also records daily occurrences of GPS jamming and spoofing in regions such as the Baltic Sea, and says regions outside conflict zones will see more electromagnetic interference (CSIS, 2025b).

The effect reaches civilians. A note to the Council of the European Union from three member states records a dramatic increase since August 2024 in jamming and spoofing of satellite navigation signals for aircraft in the Baltic region, and more frequent interference with signals received by ships since the beginning of 2025. It gives counts of cases by unique aircraft identifier per month, and describes the problem as a threat to aviation and maritime safety (Council of the EU, 2025).

Finland's transport agency said in June 2025 that disruptions in the Gulf of Finland had "already caused near misses," and that spoofing of the Automatic Identification System, which shows where ships are, had become more frequent. It added that other systems are also used to verify ship location and movement (Traficom, 2025). The EU aviation safety agency and the airline association published a joint plan in June 2025 that includes keeping a backup for satellite navigation, with a minimum operational network of traditional navigation aids (EASA, 2025).

Read the last line again. The remedy for spoofed signals is an independent second source.

Why this reaches your office. The same Council note says satellite timing services are used to synchronize telecommunication networks and financial transactions, and to stabilize power grids (Coun-

cil of the EU, 2025). Time is infrastructure. Position is infrastructure. If your systems trust a signal, someone can try to feed them a false one.

Deception works the same way against people. CSIS notes that deepfake video has already appeared in the war, and warns that the next iteration could include deepfake military orders (CSIS, 2025a). IBM found that AI deepfake or impersonation attacks made up 45% of AI-driven attacks, and that voice and SMS phishing led to the highest average breach cost among attack vectors, at USD 5.29 million (IBM and Ponemon Institute, 2026).

What it means for you. Do not trust a single signal. Verify out of band. If someone who sounds like your chief executive asks for a payment, call back on a number you already hold. If a log timestamp matters, know where your systems get their time and what happens if it is wrong. If an alert is the only evidence, find a second.

Where AI fits. AI can help cross-check signals against each other. It can also be fed bad signals. Treat everything a model reads as untrusted input, and never let one sensor, human or machine, be the only vote.

Lesson 5: Measure, countermeasure, repeat

What public analysis shows. CSIS quotes an analyst describing a "six week learning cycle on the battlefield." As new ways to jam drone signals appear, so do ways around the jamming. CSIS reports drones guided in the final phase of flight to counter jamming, and drones controlled over fiber-optic cables, which are impervious to jamming (CSIS, 2025a). Its space chapter says drone developers have pushed innovation cycles to weeks rather than months or years, and notes that jammers, when radiating, can be detected and located (CSIS, 2025b). Every fix invites the next move.

What it means for you. Controls decay. A detection that worked in spring is evaded by autumn. An AI model tuned on last year's alerts drifts as the environment changes. The joint CISA, NSA and FBI guidance on AI data security lists data drift as one of three major data risks for AI systems (CISA et al., 2025).

So schedule the countermeasure cycle. Re-test your detections. Re-run your tabletop exercises. Re-baseline your metrics. Review the AI tools quarterly and ask what the adversary has changed since the last review.

Where AI fits. AI helps you adapt faster, if you measure it. Without measurement, it is only a faster way to be wrong.

Lesson 6: The rear area is not safe

What public analysis shows. CSIS expects logistics chains to be in "constant peril," says modern forces need to anticipate that only a fraction of supplies will get through, and expects them to seek self-sufficiency (CSIS, 2025a). CSIS also says the line between combatants and noncombatants will blur, and that civilian providers of power, cloud storage and internet connectivity "could increasingly be seen as legitimate military targets" (CSIS, 2025a). That is a forecast by one think tank, not a legal

ruling. In its space chapter, CSIS notes that whether commercial satellites used for military purposes become legitimate targets is an area of active debate (CSIS, 2025b).

The same chapter gives a case of collateral reach. Early in the war, a cyberattack on a commercial satellite communications provider took tens of thousands of satellite modems offline across Central and Eastern Europe. It also knocked out wind turbines and internet access for civilians across Europe (CSIS, 2025b). The target was a commercial provider. The damage reached customers who were nowhere near the fighting.

What it means for you. Your supply chain is your rear area. Verizon found third-party involvement in 48% of breaches in its 2026 dataset (Verizon, 2026). The CISA advisory AA24-038A observed that victims included smaller organizations that provide critical services to larger ones (CISA et al., 2024).

Map the dependencies. The attack on one satellite provider reached wind turbines. Which vendors can reach your data? Which cloud, network and identity providers does the business stop without? What happens if one is unavailable for a week? Keep offline copies of what you cannot lose. Keep a manual fallback for what you cannot pause.

Where AI fits. Every AI tool is a new supply line, with a vendor, a data path and a failure mode. Add it to the map before you depend on it.

Lesson 7: Decentralize decisions, protect the network

What public analysis shows. CSIS reports that a more decentralized communications model, with multiple dispersed radio nodes and smaller radios, reduces the detectability of signals and makes them harder to jam (CSIS, 2025a). It expects future units to be "highly empowered," with field commanders making decisions about cover, concealment and tactics while minimizing their communications signature, and it says units will need resilient mesh networks so that communications continue "even if multiple nodes go dark" (CSIS, 2025a).

CSIS also names a danger. When both sides have AI-enabled processing, it warns, the incentive will lean toward deferring to the machine. "For now, that surely will lead to mistakes" (CSIS, 2025a).

What it means for you. A security operation that needs one person's approval for everything, one tool for everything or one channel for everything will break under pressure. Pre-authorize decisions. Write down who may isolate a system at 3 a.m. and under what conditions. Keep a communications path that does not depend on the systems you are defending. If your email is compromised, how does the incident team talk?

And do not build a static, automated wall and call it done. Automation that never changes is a map the adversary can study. People who can improvise, and who know the plan, are the redundancy.

Where AI fits. AI is a node. Design the operation so that losing it degrades the team and does not stop it.

Where the comparison breaks

I owe you the limits.

No one is shooting at your servers. Your adversary is mostly a criminal crew. The consequences, the law and the ethics are different. Borrowing a lesson is not borrowing a license.

The sources are analyses of an ongoing conflict, written from public data and open reporting. Facts change. The CSIS report was published in September 2025, and the other war-related documents I cite date from 2025. Forecasts are forecasts.

And I am drawing strategic lessons about technology. I am not offering tactics, and I will not.

Dead reckoning for security operations

When the signal fails, you navigate from what you know. Five parts.

1. **Last known position.** Your baseline and your inventory. What do you own, and what does normal look like?
2. **Heading.** Your priorities. The few exposures and behaviors that matter most, set in advance.
3. **Speed and time.** Your measures: time to detect, time to contain, time to fix. And a trusted source of time.
4. **A second source.** Independent cross-checks. Out-of-band verification. A backup that does not share the same failure.
5. **A fix when you can get one.** Re-verify against the best signal when it returns, and correct your course.

The rest of this paper is the detail. Part I applies the irregular warfare lens. The recommendations give you the sequence.

Analysis

The analysis has nine parts. The first two build the framing lens and apply it to adversary behavior. The rest turn that lens on the practical work of running AI inside security operations.

Part I: The irregular warfare lens

Irregular warfare is the oldest part of this story. The modern battlefield is the loudest.

Both teach the same thing. The side that waits, hides and adapts makes the other side pay.

Why bring irregular warfare into a paper about security operations?

Because the frame you use decides the questions you ask.

If you treat cyber defense purely as an IT hygiene problem, you will ask hygiene questions. Is the patch applied? Is the antivirus current? Those questions matter. But they miss the strategic logic of the people on the other side.

If you treat it as a contest with an adaptive opponent who is deliberately operating below your thresholds, you ask better questions. Where would I hide if I were them? What would make my response slow, expensive or uncertain? Which of my institutions would they try to weaken first?

That second set of questions is the irregular warfare set. It is the analytic approach I bring to every engagement, and it shapes everything that follows.

What doctrine actually says

Start with definitions, because loose language produces loose thinking.

U.S. joint doctrine distinguishes two forms of warfare. Joint Publication 1, *Joint Warfighting*, characterizes conventional warfare as "a violent struggle for domination between nation-states or coalitions and alliances of nation-states, fought with conventional forces" (CRS, 2024).

It defines irregular warfare (IW) differently:

> *"A form of warfare where states and non-state actors campaign to assure or coerce states or other groups through indirect, non-attributable, or asymmetric activities, either as the primary approach or in concert with conventional warfare." (JP 1 Vol 1, 2023, as quoted in CRS, 2024)*

Read that again and pick out the three adjectives. Indirect. Non-attributable. Asymmetric.

Every one of them describes a modern intrusion campaign.

Joint doctrine adds that the term "irregular" highlights the character of this form of warfare, which "seeks to create dilemmas and increase risk and costs to adversaries to achieve a position of advantage" (JP 1 Vol 1, 2023, as quoted in U.S. Army CALL, 2025).

Create dilemmas. Increase costs. Gain advantage. Hold onto those phrases. We will come back to them when we talk about alert fatigue, patch backlogs and board reporting.

How the definition changed

The current definition is not the old one, and the change is instructive.

According to the Congressional Research Service, previous DoD doctrine characterized IW as "a violent struggle among state and non-state actors for legitimacy and influence over the relevant population(s)" (CRS, 2024).

The older wording centered on violence and on populations, the language of insurgency and counterinsurgency. The newer wording centers on campaigning, on coercion and assurance, and on methods that are indirect, non-attributable or asymmetric.

That shift tracks a broader change in how the U.S. defense establishment talks about competition. CRS notes that the Irregular Warfare Annex to the 2018 National Defense Strategy, released by DoD in 2020, called for a revised understanding of IW to account for its role in great power competition. CRS also notes that the 2022 National Defense Strategy refers to competitors' activities in the so-called gray zone between peace and wartime (CRS, 2024).

CRS adds one more point that matters for anyone in security operations. The IW operating environment includes all domains and the information environment, and because of its emphasis on influencing populations, actions to control the information environment, including actions in cyberspace, play a prominent role in IW (CRS, 2024).

Key word: cyberspace. Doctrine already puts your network inside this contest.

The gray zone

The second concept is the gray zone.

The Center for Strategic and International Studies (CSIS), in its 2019 study *By Other Means*, describes rivals who, given the significant costs of engaging the United States in combat and the growing range of indirect and non-military tools at their disposal, are "finding avenues for threatening U.S. interests without triggering escalation." Their approaches, CSIS writes, "lie in the contested arena between routine statecraft and open warfare" (CSIS, 2019).

The CSIS study team also made a point I find refreshing. It listed the many competing labels for this space: irregular warfare, sharp power, hybrid warfare, active measures, political warfare, strategic competition, gray zone. And it noted that the lack of consensus is unsurprising, because "the strategies and operations present in the space between peace and war are intentionally ambiguous, meant to defy easy categorization and therefore also effective responses" (CSIS, 2019).

Intentionally ambiguous. That is the design goal.

RAND's 2019 study *Gaining Competitive Advantage in the Gray Zone* offers a compact working definition:

> *"The gray zone is an operational space between peace and war, involving coercive actions to change the status quo below a threshold that, in most cases, would prompt a conventional military response, often by blurring the line between military and nonmilitary actions and the attribution for events." (RAND, 2019)*

RAND also notes that gray zone aggressors aim to scale their actions to fall just short, or in some cases well short, of established triggers for a response (RAND, 2019).

Now translate that into your environment.

Your security tools are built around triggers. A threshold of failed logins. A signature match. An anomaly score. A gray-zone mindset says: find the threshold, then stay under it.

Hybrid threats

The third concept comes from NATO and the European Union: hybrid threats.

NATO describes hybrid threats as combining "military and non-military as well as covert and overt means, including disinformation, cyber attacks, economic pressure, deployment of irregular armed groups and use of regular forces." Hybrid methods, NATO says, "are used to blur the lines between war and peace, and attempt to sow doubt in the minds of target populations" (NATO, 2026).

NATO also observes something every defender should underline. Hybrid methods such as propaganda, deception and sabotage have long been used. What is new "is their speed, scale and intensity, facilitated by rapid technological change and global interconnectivity" (NATO, 2026).

Speed, scale and intensity. That is exactly what AI adds to an adversary's toolkit.

The European Union's 2016 Joint Framework on countering hybrid threats says the concept aims to capture "the mixture of coercive and subversive activity, conventional and unconventional methods (i.e. diplomatic, military, economic, technological), which can be used in a coordinated manner by state or non-state actors to achieve specific objectives while remaining below the threshold of formally declared warfare." It adds that there is usually an emphasis on "exploiting the vulnerabilities of the target and on generating ambiguity to hinder decision-making processes" (European Commission and High Representative, 2016).

Generating ambiguity to hinder decision-making. If you have ever sat in an incident bridge call where nobody could agree whether the event was real, who was behind it, or who had the authority to shut a system down, you have felt that effect.

The European Centre of Excellence for Countering Hybrid Threats (Hybrid CoE) in Helsinki characterizes hybrid threats as activities "that exploit the thresholds of detection and attribution, as well as the

different interfaces (war-peace, internal-external security, local-state, and national-international)" (Hybrid CoE).

Thresholds of detection. Interfaces between organizations. Those are SOC problems.

Six features that carry over to cyber defense

Pulling these sources together, I see six recurring features of irregular, gray-zone and hybrid competition. Each one maps directly onto the daily work of a security team. The mapping is my analysis. The features themselves come from the sources above.

1. Ambiguity and the attribution problem. Doctrine calls irregular activity non-attributable. RAND describes blurring the attribution for events. Hybrid CoE describes exploiting thresholds of attribution. In security operations, this shows up as intrusions you cannot confidently assign to a named actor, campaigns that reuse common tools, and infrastructure rented or hijacked from innocent third parties. The practical lesson: build your response around behavior and impact, not around certainty about who did it. You will rarely get certainty in time.

2. Threshold management. CSIS lists bounded thresholds as a common element of gray-zone activity: actions more threatening than statecraft but designed to avoid escalatory tripwires (CSIS, 2019). In a network, the tripwires are your detection rules. An adversary operating in this mode uses legitimate accounts, built-in tools and normal working hours. The practical lesson: thresholds alone will not find a patient intruder. You need baselines, correlation and people who know what normal looks like.

3. Influence and the information environment. CRS notes that actions to control the information environment, including cyberspace, play a prominent role in IW (CRS, 2024). NATO and the EU both emphasize disinformation. For a mid-size organization, the influence problem is closer to home than geopolitics: phishing that impersonates executives, fraudulent payment requests, fake vendor messages, and now synthetic voice and text that read as authentic. The practical lesson: your people are part of your sensor network, and your verification procedures are part of your defense.

4. Proxies and intermediaries. The EU framework mentions proxy actors. Hybrid CoE notes that non-state actors often act as proxies for state sponsors (Hybrid CoE). In cyber terms, proxies include criminal groups, compromised third-party suppliers, hijacked home and small-office routers used to relay traffic, and managed service providers whose access becomes an attacker's access. The practical lesson: your risk extends to every party with a connection into your environment.

5. Asymmetric cost imposition. Joint doctrine says irregular warfare seeks to "increase risk and costs to adversaries." Hybrid CoE describes hybrid threats as "a cost-efficient way" for authoritarian states to achieve their objectives (Hybrid CoE). In security operations, asymmetry is everywhere. An attacker sends a thousand cheap phishing messages. You spend analyst hours on each one that gets reported. An attacker scans for one unpatched appliance. You must patch all of them. The practical lesson: a defender who tries to match the attacker effort for effort loses. You need leverage, and leverage means automation, prioritization and design choices that make attacks expensive.

6. Resilience of civil institutions. This is the feature defenders most often forget, and it is the one doctrine emphasizes most. NATO's approach to hybrid threats stresses national resilience and civil preparedness (NATO, 2026). Hybrid CoE says democratic states need to "foster resilience against these threats in both government and the broader civil society," including protection of critical infrastructure (Hybrid CoE). RAND's proposed strategic concept includes sustaining resilience in the lower-level, persistent competition areas (RAND, 2019). For your organization, resilience means you can keep operating, make decisions under uncertainty, and recover quickly. It means backups that work, incident roles that are clear, and leaders who have rehearsed.

> *"You will not out-spend an irregular adversary. You can out-organize one.*

The mapping in one table

The table below summarizes how each feature appears in a security operations context, and what kind of AI support may help. It is Infinitum's practical judgment, informed by the sources above. It is not a measured result.

IW / gray-zone feature	How it shows up in your environment	Where AI may help	Where people must decide
Ambiguity and attribution	Intrusions using common tools and borrowed infrastructure	Clustering related alerts; summarizing what is known and unknown	Whether to act on incomplete information; external notification
Threshold management	Valid accounts, built-in tools, activity during business hours	Baselining normal behavior; surfacing small deviations	Whether a deviation is benign; tuning thresholds
Influence and information	Executive impersonation, payment fraud, synthetic content	Flagging suspicious messages; drafting user guidance	Verification procedures; communications to staff
Proxies and intermediaries	Supplier access, managed service accounts, relayed traffic	Inventorying third-party access; enriching external infrastructure	Supplier risk decisions; contract terms
Asymmetric cost imposition	Alert floods, patch backlogs, repeated low-cost attempts	Triage, deduplication, prioritization	Which work to stop doing; risk acceptance
Resilience	Outages, ransomware, decision paralysis	Drafting playbooks and after-action notes; exercise support	Recovery priorities; when to restore

What the counter-hybrid playbooks prescribe

Definitions describe the threat. The more useful material, for a defender, is what governments and research institutions recommend doing about it. Four sources stand out, and each translates surprisingly well to a mid-size security program.

NATO: prepare, deter, defend. NATO states that since 2015 it has had a strategy for its role in countering hybrid warfare, built around preparing, deterring and defending (NATO, 2026). To be prepared, NATO says it "continuously gathers, shares and assesses information in order to detect and attribute any ongoing hybrid activity." It supports Allies in identifying national vulnerabilities and strengthening resilience. It emphasizes training, exercises and education, including exercising decision-making processes. And since 2016, the Alliance has publicly stated that hybrid actions against one or more Allies could lead to a decision to invoke Article 5 of the North Atlantic Treaty (NATO, 2026).

Strip away the alliance politics and you have a security operations model. Gather, share and assess information. Know your own vulnerabilities. Exercise decisions, not just tools. And make clear in advance what level of activity triggers what response, so the adversary cannot exploit your hesitation.

NATO: tailored support teams. In July 2018, NATO leaders agreed to set up counter-hybrid support teams that provide tailored, targeted assistance to Allies on request (NATO, 2026). The principle is worth borrowing. The primary responsibility to respond rests with the targeted country, NATO says, but outside expertise is available when asked. A mid-size organization works the same way. You own your defense. You bring in specialized help for specific gaps.

RAND: four complementary efforts. RAND's proposed strategic concept for responding to gray zone threats rests on four complementary efforts: to shape a context supportive of U.S. and partner objectives over the long term; to deter a handful of very extreme forms of gray zone aggression; to dissuade the day-to-day use of more-elaborate gray zone techniques; and to sustain resilience in the lower-level, persistent competition areas (RAND, 2019).

I find this the most useful of the four for security leaders, because it separates problems by severity and frequency. Translated:

- **Shape:** design your architecture, identity model and supplier relationships so that the environment favors the defender over the long term.
- **Deter the extreme:** for the handful of catastrophic scenarios, such as ransomware across all systems or loss of control of operational technology, invest in strong, specific protections and rehearsed recovery.
- **Dissuade the elaborate:** make the more sophisticated techniques expensive and noisy, with baselines, logging and early warning.
- **Sustain resilience against the persistent:** accept that low-level probing, phishing and scanning will never stop, and build processes that absorb them cheaply. This is where AI-assisted triage earns its keep.

Hybrid CoE: awareness and resilience. The Helsinki centre says the first step in defending against hybrid threats is to raise awareness, and that democratic states need to "foster resilience against these threats in both government and the broader civil society," with examples including media literacy and protection of critical infrastructure (Hybrid CoE). For your organization: awareness means your staff, not just your SOC, understand that they are targets. Resilience means the organization keeps functioning when a system, a supplier or a decision-maker is unavailable.

Joint doctrine: address root causes. One more doctrinal idea is worth importing. CRS quotes JP 3-24, which describes counterinsurgency as "comprehensive civilian and military efforts designed to simultaneously defeat and contain insurgency and address its root causes" (CRS, 2024). I am not suggesting that a network intrusion resembles an insurgency. The analytic point is narrower. Effective responses to irregular threats work on two tracks at once: containing the immediate problem and fixing the conditions that allowed it.

In security operations, the immediate problem is the incident. The root causes are the exposed vulnerability, the over-privileged account, the missing log, the unclear authority. A SOC that only contains incidents will contain the same incident again. AI that only speeds up containment, without feeding root-cause fixes, makes you faster at losing slowly.

> *Contain the incident. Fix the condition. Do both, every time, or you will meet the same adversary through the same door.*

What the lens does not mean

One clarification before we move on.

Using an irregular warfare lens does not mean every intrusion is a state campaign. Most are not. Many are financially motivated crime.

It also does not mean a mid-size company should think of itself as a combatant. You are not. Your job is defense, recovery and continuity, within the law.

The value of the lens is analytic. It reminds you that your opponent is adaptive, cost-conscious and deliberately ambiguous, whatever their motive. Criminal groups borrow gray-zone habits because those habits work. Staying under thresholds, using legitimate tools and exploiting slow decision-making pay off for anyone.

The lens also keeps strategy in view. A SOC that only counts alerts is fighting at the tactical level. A SOC that asks how the organization stays resilient under persistent pressure is thinking at the level where the contest is actually decided.

Next, let's look at how adversaries adapt in practice.

Part II: Adversary adaptation

Irregular actors adapt. That is close to the definition of the term.

They adapt to your defenses, to your politics, to your budget cycle and to your tools. When you close one door, they test the next. When a technique becomes noisy, they replace it with a quieter one.

AI accelerates that loop. To see what adaptation looks like in a real, publicly documented case, let's look at a joint cybersecurity advisory from U.S. and partner agencies.

A worked example: persistence in critical infrastructure

In February 2024, CISA, the National Security Agency and the FBI, joined by partner agencies, published joint advisory AA24-038A on a persistent-access campaign against critical infrastructure networks (CISA et al., 2024).

I use this advisory as a teaching case because it is public, detailed and authoritative. I use it as a source on techniques only. I do not assess any government or group, and I leave out the advisory's attribution. I will stay at the level of strategy and defense. The advisory itself contains the technical detail defenders need.

Here is what the authoring agencies said, in summary.

They observed indications of the actors "maintaining access and footholds within some victim IT environments for at least five years" (CISA et al., 2024).

And they noted that "some victims are smaller organizations with limited cybersecurity capabilities that provide critical services to larger organizations or key geographic locations" (CISA et al., 2024).

Let that sink in. Five years. Smaller organizations. Exposed because of who and what they connect to.

Reading the case through the lens

Now apply the six features from Part I.

Thresholds. The advisory describes heavy use of living-off-the-land techniques, meaning the use of tools already built into the operating system, together with valid accounts and strong operational security, which combined "allows for long-term undiscovered persistence" (CISA et al., 2024). In some instances, the agencies note, the actors may have avoided using compromised credentials outside normal working hours to avoid triggering alerts. That is threshold management in its purest form.

Ambiguity. The advisory notes that conventional indicators of compromise are generally lacking for this kind of activity, which complicates defenders' efforts to identify, track and categorize it (CISA et al., 2024). No malware signature to match. Behavior that looks like administration.

Proxies. The advisory reports that the actors used multi-hop proxies, including virtual private servers and end-of-life small office and home office routers, to support their operations (CISA et al., 2024). Borrowed infrastructure. Somebody else's equipment.

Asymmetry. The advisory says the actors commonly gained initial access by exploiting known or zero-day vulnerabilities in public-facing network appliances, often using publicly available exploit code for known vulnerabilities (CISA et al., 2024). That is cheap for the attacker and expensive for the defender who has to find and fix every exposed device.

Resilience. The incident response guidance in the advisory includes planning for how to keep operating if access to or control of IT or operational technology is lost, and regularly testing manual controls (CISA et al., 2024). The agencies are planning for degraded operations, not just for detection.

What the defenders were told to do

The advisory's list of actions to take today is short and almost boring (CISA et al., 2024):

1. Apply patches for internet-facing systems, prioritizing critical vulnerabilities in appliances known to be frequently exploited.
2. Implement phishing-resistant multifactor authentication.
3. Ensure logging is turned on for application, access and security logs, and store logs in a central system.
4. Plan "end of life" for technology beyond the manufacturer's supported lifecycle.

Not one of those actions requires AI.

That is the first lesson of this case. Fundamentals beat novelty. An AI program built on top of unpatched edge devices, weak authentication and missing logs is decoration.

The second lesson is subtler. The advisory says many organizations "lack security and network management best practices (such as established baselines) that support detection of malicious LOTL activity," which "makes it difficult for network defenders to discern legitimate behavior from malicious behavior and conduct behavior analytics, anomaly detection, and proactive hunting" (CISA et al., 2024).

That is precisely the kind of problem where AI-assisted analysis may help, if, and only if, the logs exist and someone has defined what normal looks like.

Sharing as a defensive multiplier

One more lesson from the advisory is easy to overlook. Its incident response guidance includes reporting compromises to the authoring agencies and considering sharing technical information with an authoring agency or a sector-specific information sharing and analysis center (CISA et al., 2024).

That matters for the irregular warfare lens. An adversary operating across many victims sees patterns that no single victim can see. Each defender, alone, sees one fragment. Sharing reverses that asymmetry. When defenders pool what they observe, the adversary's breadth becomes a liability rather than an advantage, because the same behavior seen in many places becomes easier to recognize everywhere.

NATO makes the same point at the alliance level, describing how it "continuously gathers, shares and assesses information in order to detect and attribute any ongoing hybrid activity" (NATO, 2026).

For a mid-size organization, sharing does not require a classified relationship. It can be as simple as membership in a sector information-sharing group, a relationship with a regional law-enforcement contact, and a habit of reading and acting on public advisories. AI can help here too. Summarizing a long advisory into the three checks your team must run this week is exactly the kind of task it handles well, as long as a person verifies the summary against the original.

Watching your own adaptation

Adaptation runs both ways, and defenders are often slower than they think.

A useful exercise, borrowed from the habit of after-action review in irregular warfare analysis, is to ask after every significant incident or exercise: what did the adversary do that we did not expect, and what will we change because of it? If the honest answer is "nothing," you are not adapting. You are repeating.

Write down the answer. Assign an owner. Check the change was made. That loop, run consistently, is worth more than any single tool.

How AI changes the adaptation loop

Now add AI to the adversary side.

Verizon reports that AI is helping threat actors speed up the time it takes to exploit known vulnerabilities (Verizon, 2026). IBM reports that more than one in four organizations hit by a malicious attack said it was AI-driven, and that AI-driven attacks added an average of USD 1 million to the cost of a malicious breach (IBM and Ponemon Institute, 2026).

From an irregular warfare perspective, AI does three things for an adversary.

It lowers the cost of scale. Tasks that once took skilled time, such as writing convincing messages in a target's language or sorting stolen data, become cheaper. That widens the asymmetry between attacker cost and defender cost.

It shortens the cycle. The gap between disclosure of a vulnerability and its exploitation shrinks. Your remediation calendar, measured in weeks, now competes with an attacker cycle that is getting shorter.

It improves the disguise. Influence operations and impersonation become more convincing. The information-environment feature of irregular competition reaches your inbox and your phone.

None of this changes the strategic logic. It intensifies it. NATO's observation about hybrid threats applies here too: the methods are old; the speed, scale and intensity are new (NATO, 2026).

How defenders should adapt

If the adversary adapts, so must you. But adapt deliberately.

The irregular warfare tradition offers a useful discipline here. Effective counter-irregular strategies do not try to match the opponent everywhere. They protect what matters most, deny the opponent easy wins, build the resilience of the institutions under pressure, and learn faster than the opponent.

Translated to a mid-size security program, that means four things.

1. **Protect what matters most.** Know your critical assets and data flows. You cannot defend everything equally.
2. **Deny easy wins.** Fix the exposed, known-exploited vulnerabilities first. Lock down authentication. Turn on logging.
3. **Build institutional resilience.** Rehearse decisions. Test backups. Clarify authority.
4. **Learn faster.** Measure what happens. Review AI outputs and analyst overrides. Feed every incident back into detection and design.

AI can help with parts of all four. It cannot do any of them for you.

Part III: The evidence on AI for defenders

So does AI actually help defenders? The evidence says yes, with an asterisk.

The good news

IBM compared breached organizations by their use of security AI and automation. Those using it extensively averaged USD 4.00 million per breach. Those using none averaged USD 5.93 million. That is a USD 1.93 million difference. Extensive users also identified and contained breaches in 215 days on average, against 280 days. Limited users landed in between, at USD 5.05 million and 244 days (IBM and Ponemon Institute, 2026).

Use of security AI and automation	Average breach cost (USD)	Mean time to identify and contain
Extensive	4.00 million	215 days
Limited	5.05 million	244 days
None	5.93 million	280 days

Source: IBM and Ponemon Institute, Cost of a Data Breach Report 2026.

The asterisk

Here's the asterisk. This is an association in a benchmark study, not proof that a tool will do the same for you.

Organizations that invest in automation may simply be better run in other ways. They may have bigger budgets, more mature processes or better-trained staff. The study cannot fully separate those effects.

Still, the direction is hard to ignore: faster, better-prioritized work gives attackers less time inside your environment.

From the irregular warfare perspective, time is the decisive variable. A patient intruder needs time to learn, move and position. Every day you take away from that timeline is a day of advantage taken from them.

The practitioner view

ISC2 points the same way from the practitioner side. 45% of respondents said their cybersecurity teams were already using generative AI in their tools, mostly to augment common operational tasks, write reports and support incident reporting. They also expected the biggest return on investment to take two or more years (ISC2, 2024).

Plan for a gradual rollout. Anyone promising transformation by next quarter is selling something.

The uneven rollout

Only 36% of breached organizations used security AI and automation extensively. 25% used none. Among organizations with a SOC that deployed AI agents, the most common uses were threat hunting (56%) and automated response and containment (54%). Only 18% pointed agents at vulnerability scanning and management (IBM and Ponemon Institute, 2026).

Let that sink in.

Verizon says vulnerability exploitation is the leading way in. Yet vulnerability management is where AI agents are least used. Detection is well covered. Exposure reduction is where many teams have the most room to improve.

In irregular warfare terms, many organizations are investing in watching the perimeter and reacting to contact. Fewer are investing in denying the opponent the cheap, easy access route in the first place. That is an imbalance worth correcting.

The new risks

This is not a one-way street.

IBM reports that 21% of organizations had a security incident involving an AI model or application, up from 13% the year before. The most common causes were compromised connected apps, APIs or plug-ins (27%) and cloud misconfigurations affecting AI workloads (27%). Among organizations with an AI-related breach, 92% lacked proper AI access controls. In IBM's words, these were governance failures, not model risk (IBM and Ponemon Institute, 2026).

Then there is shadow AI, meaning employees using unapproved AI tools. IBM reports that the share of security incidents involving shadow AI more than doubled to 43% from 20%, and that 68% of breached organizations lacked governance to manage AI or detect shadow AI. Verizon reports that frequent use of AI tools by employees rose from 15% to 45% in a year (Verizon, 2026). ISC2 found that 54% of respondents had already faced data privacy and security concerns from their organizations' adoption of generative AI (ISC2, 2024).

This is not an AI problem. It is a basics problem, with AI on top.

Access control. Asset inventory. Approved tools. Configuration management. The same fundamentals that the advisory AA24-038A emphasizes for networks apply, almost word for word, to AI systems.

Hold that thought. We return to it in Parts V and VI.

Part IV: The use cases, one by one

Not every task suits AI equally.

This part walks through six common security operations use cases. For each, I cover what the work is, what AI can plausibly contribute, what can go wrong, how a person stays in control, and what to measure. I close each with the irregular warfare angle, because it often changes the priority.

First, the summary table. It is Infinitum's practical judgment, informed by the sources in this paper. It is not a measured result.

Use case	Potential benefit	Main risk	Human oversight
Alert triage	Faster first-pass summaries and grouping of related alerts	Missed or wrongly dismissed true alerts	Analyst confirms every closure; sample reviews of AI decisions
Threat-intel enrichment	Context on indicators and techniques, mapped to MITRE ATT&CK	Unverified or outdated claims presented as fact	Analyst checks sources before action
Vulnerability prioritization	Ranks fixes using CISA KEV status, exposure and asset value	Over-trusting scores; skipping context	Owners approve patch order and exceptions
Log analysis	Plain-language search and pattern spotting across large volumes	Sensitive data sent to unapproved tools	Approved tools only; engineers validate queries
Reporting and documentation	Drafts incident notes and status reports faster	Inaccurate or overconfident statements	Named owner edits and signs off
Deception and early warning	Helps plan, monitor and summarize decoy activity	Decoys that leak, mislead staff or create legal exposure	Written approval, legal review, tight scope

1. Alert triage

The work. Security tools generate alerts. Someone has to look at each one, decide whether it matters, group it with related alerts, and either close it or escalate it. In a mid-size organization, that someone is often one or two people, sometimes part-time.

What AI can contribute. A first-pass summary of each alert in plain language. Grouping of related alerts into a single case. A suggested severity with the reasoning shown. A draft of the questions an

analyst should answer next. Pulling in context, such as who owns the affected system and whether the user normally does this.

What can go wrong. The serious failure is a true alert wrongly dismissed. A model that is right most of the time can still be wrong on the alert that matters. Worse, an analyst who has watched the model be right a hundred times stops checking. That is automation bias, and it is the quiet killer of AI-assisted triage.

The second failure is noise disguised as help: long, fluent summaries that take as long to read as the raw alert.

How a person stays in control. AI proposes. A person decides. No alert is closed without a human confirmation, at least during the pilot. Keep a record of every AI suggestion and every analyst override. Each week, pull a sample of AI-recommended closures and have a second analyst review them blind.

What to measure. Time from alert to first human decision. Share of AI recommendations that analysts overrode, and why. Number of true incidents that were first suggested for closure. Analyst time per alert before and after.

The irregular warfare angle. Alert floods are a form of asymmetric cost imposition, whether intended or not. Every cheap attempt by an attacker costs you analyst minutes. Triage support is where you recover that asymmetry. But remember threshold management: the patient intruder generates few alerts, not many. Do not let triage efficiency become the only goal. A faster queue is worthless if the dangerous event never enters it.

2. Threat-intelligence enrichment

The work. An alert contains an IP address, a domain, a file hash or a suspicious behavior. Enrichment means finding out what is known about it. Has it been seen before? Is it associated with a known campaign? Which adversary techniques does it suggest?

What AI can contribute. Rapid summarization of what internal and approved external sources say about an indicator. Mapping observed behavior to MITRE ATT&CK, the knowledge base of adversary tactics and techniques that helps you map detections and find coverage gaps (MITRE ATT&CK). Drafting a short context note for the analyst. Translating a dense public advisory into the three things your team needs to check today.

What can go wrong. Unverified or outdated claims presented as fact. A model may connect an indicator to a named group on thin evidence, or repeat an association that was true years ago. Shared infrastructure makes this worse: an address used by an attacker last year may belong to an innocent tenant today.

How a person stays in control. Every enrichment claim cites its source, and the analyst checks the source before acting. Attribution statements are flagged as low confidence by default. Nobody takes

action against an external party, and nobody makes an external attribution statement, on the strength of an AI-generated summary.

What to measure. Time to context for a new indicator. Share of enrichment claims that could be traced to a source. Number of corrections made by analysts.

The irregular warfare angle. This is where the attribution problem lives. Doctrine describes irregular activity as non-attributable, and RAND describes the blurring of attribution as a defining gray-zone tactic (RAND, 2019). An AI tool that produces confident attribution is producing exactly the false certainty that ambiguity is designed to exploit. Use enrichment to understand behavior and likely impact. Treat the question of who as secondary, slow and best left to people and to authorities with the evidence to answer it.

3. Vulnerability prioritization

The work. Scanners find more vulnerabilities than any team can fix. Someone has to decide what gets fixed first, by whom, and by when, and track whether it actually happened.

What AI can contribute. Ranking fixes by combining several signals: whether the vulnerability appears in the CISA KEV catalog, whether the affected system is reachable from the internet, how important the asset is to the business, and whether a compensating control exists. Grouping findings by system owner so that each owner gets one clear list. Drafting change tickets. Mapping findings to ATT&CK techniques so you can see which adversary behaviors a fix would deny. Flagging technology that is past its supported lifecycle.

What can go wrong. Over-trusting a score. Skipping context that only the system owner knows, such as a business process that cannot tolerate a reboot this week. Treating the ranked list as the decision rather than as an input to it.

How a person stays in control. System owners approve patch order and exceptions. Every exception has a named owner, a reason and an expiry date. Security sets the policy; owners execute; leadership accepts residual risk in writing.

What to measure. Time to remediate KEV-listed vulnerabilities on internet-facing systems. Number of open KEV vulnerabilities. Share of exceptions past expiry. Compare your numbers with your own baseline, not with a survey median.

The irregular warfare angle. This is the use case I would move to the front of the line for most mid-size organizations, and the evidence supports that view. Verizon finds vulnerability exploitation is now the leading way in, at 31% of breaches (Verizon, 2026). Advisory AA24-038A describes initial access through vulnerabilities in public-facing appliances (CISA et al., 2024). Yet IBM finds only 18% of organizations with SOC agents pointed them at vulnerability management (IBM and Ponemon Institute, 2026).

Denying the opponent the easy route is the most cost-effective move in any irregular contest. Every known-exploited vulnerability you close on an internet-facing system removes a cheap option from the attacker's menu. That is asymmetric cost imposition in your favor.

> *Detection tells you the adversary is inside. Exposure reduction keeps them out. Most teams have the ratio backward.*

4. Log analysis

The work. Logs record what systems and people did. They are the evidence base for detection, investigation and hunting. They are also enormous, inconsistent and hard to search without specialized query skills.

What AI can contribute. Turning a plain-language question into a draft query. Summarizing patterns across large volumes. Highlighting deviations from a baseline: an account active at an unusual hour, a built-in administrative tool used on a server where it is never used, a new connection between two systems that do not normally talk.

What can go wrong. Two big risks. First, sensitive data sent to unapproved tools. Logs contain usernames, internal addresses, sometimes personal data. Pasting them into a public AI service is a data leak. Second, a plausible but wrong query that silently misses what you were looking for. A query that returns nothing feels reassuring. It may simply be wrong.

How a person stays in control. Approved tools only, with a plain rule about what data may go into them. Engineers validate any AI-drafted query before its results are relied on, and test it against a known example. Keep AI-generated detection logic in version control like any other code.

What to measure. Time to answer an investigative question. Number of AI-drafted queries that needed correction. Coverage: share of critical systems sending logs to the central store.

The irregular warfare angle. This is the use case that most directly confronts threshold management. Advisory AA24-038A makes the point that defenders who lack established baselines find it difficult to discern legitimate from malicious behavior (CISA et al., 2024). AI-assisted analysis may help build and watch those baselines at a scale a small team cannot manage by hand. But it depends entirely on the logs existing, being centralized and being protected. The same advisory recommends turning logging on and storing logs centrally. Do that first.

5. Reporting and documentation

The work. Incident notes. Shift handovers. Status reports. Board updates. Regulatory and contractual notifications. After-action reviews. In a stretched team, writing is often the work that slips, and that slippage has real costs: lost context, weak lessons learned, unclear decisions.

What AI can contribute. First drafts, faster. Turning a timeline of analyst notes into a readable narrative. Producing different versions of the same facts for different audiences: technical detail for engineers, impact and decisions for executives. This is one of the uses ISC2 respondents most commonly reported (ISC2, 2024).

What can go wrong. Inaccurate or overconfident statements. A draft that says the incident is contained when it is not. A report that smooths over uncertainty because fluent prose sounds certain. In a regulatory notification, that can create legal exposure.

How a person stays in control. A named owner edits and signs off every external or executive document. Drafts mark facts, estimates and unknowns differently. Nothing leaves the organization without human review.

What to measure. Analyst hours spent on reporting, before and after. Time from incident closure to completed after-action review. Number of factual corrections made to AI drafts.

The irregular warfare angle. This is where resilience of institutions becomes concrete. The EU framework notes that hybrid threats aim at "generating ambiguity to hinder decision-making processes" (European Commission and High Representative, 2016). Clear, timely, honest reporting is the counter. Leaders who receive accurate information quickly make better decisions under pressure. A reporting process that is fast but unreliable feeds the very ambiguity an adversary wants.

6. Deception and early warning

The work. Deception, in defensive security, means placing decoy assets, credentials or data inside your own environment so that an intruder who touches them reveals themselves. Early warning more broadly means arranging your defenses so that you learn about an intrusion early in its life rather than late.

MITRE describes its Engage framework as one "for planning and discussing adversary engagement operations." It makes a point worth repeating: "Deception is a process, not a fire-and-forget technology stack." And it frames the logic this way: with traditional cyber defense, "the adversary only needs to be right once," but with adversary engagement, "the adversary only needs to be wrong once" (MITRE Engage).

What AI can contribute. Helping plan where decoys would be most informative given your network layout. Monitoring decoy interactions and summarizing them for an analyst. Drafting realistic but fake content so decoys are believable. Correlating a decoy touch with other activity to help an analyst judge whether it is a real intrusion or a misconfigured script.

What can go wrong. Decoys that leak into production use or confuse legitimate staff. Decoy data that looks too much like real personal or customer data. Activity that drifts from passive detection toward something that could be seen as interacting with an outside party. Each of these creates operational and legal risk.

How a person stays in control. Written approval from leadership. Legal review before deployment. Decoys stay strictly inside your own environment. No decoy ever contains real sensitive data. The program is passive and defensive: it detects and informs. It never reaches outward.

What to measure. Number of decoy alerts, and the share that were confirmed as real intrusions versus internal noise. Time from decoy touch to analyst decision.

The irregular warfare angle. Deception has a long history in competition below the threshold of war. NATO lists deception among the hybrid methods long used to destabilize adversaries (NATO, 2026). In a defensive setting, the logic flips. You use the adversary's need to explore your environment against them. A patient intruder who stays under every threshold still has to look around. A well-placed decoy turns that exploration into an early warning.

This is also the most specialized use case on the list. For most mid-size organizations, I would sequence it after the basics and after the first four use cases are working. It is a force multiplier, not a foundation.

Questions to ask before you buy

Most mid-size organizations will not build AI tools. They will buy them, or switch on AI features inside products they already own. Either way, the decision deserves the same scrutiny as any other system that touches sensitive data.

Here are the questions I would put to any vendor, in writing, before a pilot. They follow directly from the risks in this paper and from the frameworks discussed in Parts V and VI.

1. **Data.** What data does the tool send outside our environment? Where is it stored, for how long, and is it used to train models that serve other customers?
2. **Access.** What permissions does the tool need? Can it run in read-only or recommend-only mode? Can we restrict it to specific systems?
3. **Identity.** Does it support our single sign-on, multifactor authentication and role-based access? How are its service accounts and keys managed?
4. **Logging.** Does it log its own inputs, outputs and actions in a form we can send to our central log store?
5. **Off switch.** How quickly can we disable it, and what happens to in-flight actions when we do?
6. **Untrusted content.** How does the tool handle hostile input, such as phishing emails, malicious web pages or attacker-written log entries? What does the vendor do about prompt injection?
7. **Supply chain.** Which third-party models, plug-ins and services does it rely on, and how are they secured and updated?
8. **Evidence.** What independent testing has the product undergone? Will the vendor allow it to be included in our own authorized security testing?
9. **Errors.** How does the tool show uncertainty? How do analysts flag and correct wrong output, and does that feedback improve the product without exposing our data?

10. **Exit.** If we stop using it, how do we get our data back and confirm it has been deleted?

A vendor who cannot answer these clearly is telling you something. Listen.

Sequencing the use cases

If you take one thing from this part, take the sequence.

1. Start with **reporting and documentation** and **alert triage summaries**. They are low-risk, quick to measure and build trust.
2. Move fast to **vulnerability prioritization**. It addresses the leading way in and is under-used.
3. Add **log analysis** once logging coverage and data rules are in place.
4. Use **threat-intel enrichment** throughout, with strict source discipline.
5. Consider **deception and early warning** once the foundation holds.

That sequence is a judgment, not a law. Your environment may justify a different order. But be suspicious of any plan that starts with autonomous response and ends with patching.

Part V: Governance and AI risk

You don't have to invent the rulebook.

Public frameworks give you a common language, and they are free. The challenge is not finding guidance. It is applying a small, sensible slice of it consistently.

The NIST AI Risk Management Framework

The NIST AI Risk Management Framework (AI RMF) was released on January 26, 2023. It is intended for voluntary use and to improve the ability to incorporate trustworthiness considerations into the design, development, use and evaluation of AI products, services and systems (NIST, AI RMF).

Its core is organized around four functions: govern, map, measure and manage. NIST describes governance as a cross-cutting function "infused throughout AI risk management" that enables the other three (NIST, AI RMF Core).

Here is how I translate those four functions for a security team adopting AI tools. The functions are NIST's. The translation is mine.

Govern. Who decides, under what rules? NIST's govern function includes outcomes such as mechanisms to inventory AI systems, clear roles and responsibilities, and policies that address risks from third-party software and data (NIST, AI RMF Core). For you, that means: one accountable owner for AI in security, an approved-tools list, a plain rule on what data may go into AI services, and a register of every AI-enabled tool and what it can touch.

Map. What is this tool for, and what could go wrong? NIST's map function establishes the context to frame risks, including documenting the system's knowledge limits and how its output will be overseen

by humans, and defining processes for human oversight (NIST, AI RMF Core). For you: write one page per AI use case stating its purpose, its inputs, its outputs, who reviews them and what it must never do.

Measure. Is it working, and is it safe? NIST calls for AI systems to be tested before deployment and regularly while in operation, and for security and resilience to be evaluated and documented (NIST, AI RMF Core). For you: baseline metrics before the pilot, a log of AI outputs and analyst overrides, and periodic review.

Manage. What do we do about what we learned? NIST's manage function includes mechanisms "to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use," and monitoring of pre-trained models used in development (NIST, AI RMF Core). For you: a documented way to switch the tool off, a named person with authority to do it, and a review of every incident involving the tool.

Key word: deactivate. If you cannot turn it off cleanly, you do not control it.

The Generative AI Profile

On July 26, 2024, NIST released NIST AI 600-1, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST says the profile can help organizations identify unique risks posed by generative AI and proposes actions for generative AI risk management (NIST, AI RMF).

Most of the AI tools a mid-size security team will touch first are generative: summarizers, assistants, report drafters, query helpers. The profile is the right companion document for those.

NIST also notes on the same page that the AI RMF 1.0 is being revised, and that in April 2026 it released a concept note for an AI RMF profile on trustworthy AI in critical infrastructure (NIST, AI RMF). If you operate in or supply critical infrastructure, watch that work.

The NIST Cybersecurity Framework 2.0

The NIST Cybersecurity Framework (CSF) 2.0 was released on February 26, 2024. Its core is organized around six functions: Govern, Identify, Protect, Detect, Respond and Recover. Govern was newly added in 2.0 (NIST, 2024).

The CSF gives you a structure for managing risk across the whole program. AI tools are not a separate program. They sit inside it.

A practical way to use the two frameworks together:

CSF 2.0 function	AI-in-security question to answer
Govern	Who owns AI risk in security? Is there an approved-tools list and a data rule?
Identify	Do we have an inventory of AI tools, their service accounts and their data access?
Protect	Do AI tools and the data behind them have role-based access and multifactor authentication?

CSF 2.0 function	AI-in-security question to answer
Detect	Are AI tool activity and AI-generated actions logged and monitored?
Respond	Can we disable or roll back an AI tool quickly during an incident?
Recover	Do we review AI involvement in every incident and update controls afterwards?

Why governance comes first

IBM's numbers make the case on their own. Among organizations with an AI-related breach, 92% lacked proper AI access controls. 68% of breached organizations lacked governance to manage AI or detect shadow AI (IBM and Ponemon Institute, 2026).

Those are not sophisticated failures. They are the absence of basic rules.

From the irregular warfare perspective, ungoverned AI is an unmanaged interface, and Hybrid CoE describes hybrid threats as activities that exploit "the different interfaces" between organizations and domains (Hybrid CoE). Every unapproved AI tool an employee uses is a new interface between your data and someone else's system. An adversary does not need to defeat your controls if your data is already flowing around them.

A governance minimum for mid-size organizations

If you do nothing else in this part, do these six things. They are my recommendation, aligned with the frameworks above.

1. **Name one accountable owner** for AI in security operations, with authority to approve, pause and retire tools.
2. **Publish an approved-tools list** and a one-page rule on what data may and may not go into AI services.
3. **Inventory every AI-enabled tool**, including features quietly switched on inside existing products, with its service accounts, keys and data access.
4. **Put role-based access and multifactor authentication** on AI tools and the data behind them.
5. **Write a one-page use-case card** for each AI use: purpose, inputs, outputs, human reviewer, prohibited actions, off switch.
6. **Review quarterly**, and after every incident involving an AI tool.

That is not bureaucracy. That is the minimum you need to answer the question every board will eventually ask: what is AI doing in our security program, and who is accountable for it?

Part VI: Securing the AI itself

An AI tool in your security operations is not just a helper. It is a system with access, inputs, outputs and dependencies. It can be attacked like any other system, and in some new ways too.

This part describes the main risk categories at a level suited to decision-makers. It is defensive by design. It explains what can go wrong and how to reduce the risk. It does not describe how to carry out attacks.

Three public references

Three public resources give you the vocabulary.

OWASP Top 10 for LLM Applications (2025). The OWASP Gen AI Security Project publishes a list of the top risks for large language model and generative AI applications. The 2025 list includes Prompt Injection, Sensitive Information Disclosure, Supply Chain, Data and Model Poisoning, Improper Output Handling, Excessive Agency, System Prompt Leakage, Vector and Embedding Weaknesses, Misinformation, and Unbounded Consumption (OWASP, 2025).

MITRE ATLAS. ATLAS is MITRE's knowledge base of adversary tactics and techniques against AI-enabled systems. Its September 2026 data release lists 16 tactics, 120 techniques, 88 sub-techniques, 40 mitigations and 73 case studies (MITRE ATLAS, 2026). If your team already uses ATT&CK, ATLAS will feel familiar.

NIST AI 100-2 E2025. NIST's report *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations,* published in March 2025, provides a taxonomy and common terminology for attacks on AI systems and corresponding mitigations (NIST, 2025).

And for organizations deploying AI systems built by others, which is nearly every mid-size organization, the joint guidance *Deploying AI Systems Securely*, published in April 2024 by the NSA's Artificial Intelligence Security Center with CISA, the FBI and allied partners, provides best practices aimed at improving the confidentiality, integrity and availability of AI systems and at protecting, detecting and responding to malicious activity against AI systems and related data and services (NSA AISC et al., 2024).

Risk 1: Prompt injection

What it is. OWASP lists prompt injection as the first risk on its 2025 list (OWASP, 2025). In plain terms, it is when content fed to an AI system alters its behavior in ways its operators did not intend. The content can come from a user directly, or it can be hidden inside material the AI is asked to process, such as an email, a web page, a document or a log entry.

Why it matters in a SOC. Security AI tools read hostile content by design. That is their job. A triage assistant summarizes a phishing email. An enrichment tool reads a malicious web page. A log assistant reads whatever an attacker managed to write into a log. Every one of those inputs is potentially adversary-controlled.

How to reduce the risk. Treat all content processed by AI as untrusted. Limit what the AI tool is allowed to do based on its output: summarizing is low risk; executing changes is high risk. Keep a person between AI output and any consequential action. Separate tools that read hostile content from tools that hold privileged access.

Risk 2: Data and model poisoning

What it is. OWASP describes data poisoning as occurring when pre-training, fine-tuning or embedding data is manipulated (OWASP, 2025). The effect can be a model that behaves normally in most cases and wrongly in specific ones.

Why it matters in a SOC. If a tool learns from your own alert history or analyst feedback, an adversary with long-term access could, in principle, influence what it learns. An intruder who is present for years, as advisory AA24-038A describes, has time to shape what normal looks like.

How to reduce the risk. Know where your tools' training and reference data come from. Control who can add to it. Keep versioned copies so you can roll back. Test tools against known examples after every update.

Risk 3: The model and software supply chain

What it is. OWASP lists supply chain as a top risk (OWASP, 2025). AI tools depend on models, libraries, plug-ins, connectors and hosted services from third parties. NIST's govern and manage functions both address risks from third-party software, data and pre-trained models (NIST, AI RMF Core).

Why it matters. IBM's most common root causes of AI-related incidents were compromised connected apps, APIs or plug-ins and cloud misconfigurations affecting AI workloads, each at 27% (IBM and Ponemon Institute, 2026). That is a supply chain and configuration problem, not a model problem.

How to reduce the risk. Review the APIs, plug-ins and cloud settings around every AI workload. Inventory the service accounts and keys your AI tools use, and give each one only the access it needs. Ask vendors how they secure their models and dependencies, and get the answer in writing.

From the irregular warfare lens, this is the proxy problem again. A compromised supplier is an intermediary. Its access becomes the adversary's access.

Risk 4: Over-reliance and excessive agency

What it is. Two related risks. OWASP lists excessive agency, meaning an AI system granted more ability to act than it needs, and misinformation, meaning output that is wrong but presented credibly (OWASP, 2025). The human counterpart is over-reliance: people trusting AI output without checking it.

Why it matters in a SOC. IBM found that among organizations with a SOC that deployed AI agents, 54% used them for automated response and containment (IBM and Ponemon Institute, 2026). Automated response can be valuable. It is also exactly where excessive agency hurts most. A tool that can disable accounts, isolate machines or change firewall rules can cause an outage as easily as it can stop an intrusion.

How to reduce the risk. Grant the minimum ability needed. Start every agent in recommend-only mode. Widen its authority only when tests and metrics show it behaves as intended, and keep an easy way for a person to step in. NIST's manage function calls for mechanisms to supersede, disengage or

deactivate AI systems that perform inconsistently with intended use (NIST, AI RMF Core). Build that switch before you need it.

Risk 5: Sensitive information disclosure and shadow AI

What it is. OWASP lists sensitive information disclosure as a top risk (OWASP, 2025). In practice, much of the disclosure risk comes not from sanctioned tools but from staff using unapproved ones. IBM reports that the share of incidents involving shadow AI more than doubled to 43% (IBM and Ponemon Institute, 2026).

How to reduce the risk. Give people an approved alternative that is good enough to use. Prohibition without an alternative drives usage underground. Pair the approved-tools list with clear data rules and with monitoring for unsanctioned AI services.

Securing the AI: a short checklist

- Every AI tool is in the inventory, with an owner.
- Every AI tool's service accounts and keys are inventoried and least-privilege.
- Every AI tool that reads untrusted content is separated from tools that can take privileged action.
- Every consequential AI-suggested action passes through a person until evidence supports otherwise.
- Every AI tool can be switched off quickly by a named person.
- Every AI tool is included in authorized security testing.

That last point matters. IBM lists offensive security testing among factors associated with lower breach costs (IBM and Ponemon Institute, 2026). In our terms, that means authorized, scoped, written-permission testing of your own systems, including the AI-enabled workflows. Nothing else.

Part VII: People and the workforce

Technology does not defend anything by itself. People do.

The workforce reality

Recall the ISC2 numbers. 67% of respondents reported a staffing shortage. 90% reported one or more skills gaps. Among organizations with skills gaps, 22% reported a material breach, compared with 17% of all respondents who shared that information (ISC2, 2024).

And 45% said their teams were already using generative AI in their tools, mostly to augment common operational tasks, write reports and support incident reporting (ISC2, 2024).

So AI is already in the hands of a stretched workforce. The question is whether it is making that workforce stronger or just busier.

What changes for analysts

When AI enters a security team, the analyst's job changes in three ways.

From producing to reviewing. Less time writing the first draft of a report, a query or an alert summary. More time judging whether a draft is right. Reviewing well is a skill. It has to be taught.

From reading everything to deciding what matters. Triage support reduces the reading load. That frees time for the judgment calls: is this real, does it matter, what do we do now.

From operator to supervisor. As automation widens, analysts become supervisors of systems that act. Supervision requires understanding what the system can do, where it fails, and when to step in.

The over-reliance trap

The greatest people risk in AI-assisted security is not job loss. It is skill loss.

If junior analysts never investigate an alert from scratch because the assistant always does the first pass, they never build the instincts that let them spot when the assistant is wrong. Five years later, you have a team that can supervise a tool but cannot do the work without it.

That is a resilience problem. And resilience, as Part I showed, is the core of countering irregular pressure.

What I recommend for people

These are my recommendations, informed by the sources in this paper.

1. **Keep manual skills alive.** Have analysts regularly work cases without AI assistance, and compare their conclusions with the tool's.
2. **Teach review as a skill.** Train analysts to check sources, test queries against known examples and challenge confident language.
3. **Make overrides normal.** An analyst who overrides the AI should be thanked, not questioned. Overrides are your best quality data.
4. **Train for the AI-specific risks.** Every analyst should understand prompt injection, over-reliance and data rules in plain terms.
5. **Rehearse decisions, not just tools.** Run tabletop exercises in which the AI tool is wrong, unavailable or compromised. See whether the team can still function.
6. **Extend awareness beyond the SOC.** Influence operations target everyone. Executives, finance staff and assistants are on the front line of impersonation and fraud. Give them simple verification procedures and the authority to use them.

Building the bench

Mid-size organizations rarely win bidding wars for experienced security talent. So build it.

Look inside first. System administrators, network engineers and help-desk staff already know your environment, your users and your quirks. That knowledge is exactly what an analyst needs to judge whether an AI-flagged deviation is normal. Pair one of them with your security lead for a few hours a week, and you have started a pipeline.

Use AI as a tutor, not a crutch. An assistant that explains why an alert matters, what an ATT&CK technique means or how a query works can shorten the learning curve. The rule from earlier still applies: the learner must also do the work without it.

Write things down. Every playbook, use-case card and after-action review is training material for the next person. A team that documents well survives turnover. A team that keeps everything in one person's head does not.

And bring in outside help deliberately, for specific gaps and specific periods, with a plan to transfer the knowledge in-house. That is the same principle NATO applies with its tailored support teams: the owner keeps responsibility; specialists help on request.

Who does what

Clarity of authority is a resilience measure in its own right. Ambiguity about who decides is exactly what hybrid pressure exploits. The table below is a simple allocation of responsibilities I recommend for a mid-size organization adopting AI in security operations. Adjust titles to your structure. Keep the principle: every decision has one owner.

Role	Responsibilities for AI in security operations
Executive sponsor	Accepts residual risk in writing; approves budget; receives quarterly metrics; owns the decision to widen automation
Accountable owner for AI in security	Maintains approved-tools list, inventory and use-case cards; approves pilots; holds the off switch
Security operations lead	Runs pilots; reviews AI quality weekly; manages analyst training and manual-skill exercises
Analysts	Review every AI recommendation; record overrides and reasons; escalate anything uncertain
IT and system owners	Approve patch order and exceptions; implement fixes; maintain logging on their systems
Legal and privacy	Review data rules, vendor terms, decoy programs and external notifications
All staff	Use approved tools only; follow verification procedures for payment and identity requests; report suspicious messages

A note on leadership

Leaders set the tone. NIST's govern function includes executive leadership taking responsibility for decisions about risks associated with AI system deployment (NIST, AI RMF Core).

That is not a delegation. A leader who says “the security team handles AI” has not governed it.

Ask your security lead three questions every quarter. What is AI doing in our security program? What did it get wrong? Who checked?

Part VIII: Metrics that matter

No baseline, no proof. It is that simple.

Before you pilot anything, write down today’s numbers. Then measure the same way after. Without that, every claim of improvement is an opinion.

Principles

Measure outcomes, not activity. The number of alerts processed is activity. The time a real intrusion went unnoticed is an outcome.

Measure quality alongside speed. Faster triage that misses true incidents is a regression, not an improvement.

Measure the AI, not just the team. Track how often analysts override the tool, and why.

Compare with yourself. Survey medians, including the ones in this paper, are context. Your own baseline is the benchmark.

A starter metric set

The table below is a starter set I recommend for mid-size organizations. Choose a handful. Measure them consistently.

Area	Metric	Why it matters
Exposure	Median days to remediate KEV-listed vulnerabilities on internet-facing systems	Exploitation is the leading way in (Verizon, 2026)
Exposure	Count of open KEV-listed vulnerabilities	Direct measure of easy routes left open
Exposure	Share of technology beyond supported lifecycle, with a plan	End-of-life systems are a recurring weakness (CISA et al., 2024)
Detection	Share of critical systems sending logs to a central store	No logs, no detection
Detection	Time from alert to first human decision	Speed of the first look
Response	Time to identify and contain confirmed incidents	The outcome IBM measures across its sample
AI quality	Analyst override rate, with reasons	Shows where the tool is wrong or unclear
AI quality	True incidents first suggested for closure	The single most important AI failure metric

Area	Metric	Why it matters
AI governance	Share of AI tools in inventory with an owner and use-case card	Governance coverage
AI governance	Detected use of unapproved AI services	Shadow AI exposure
People	Analyst hours spent on reporting	Where AI drafting should help
People	Exercises completed with an AI-failure injection	Resilience of the team

Reporting metrics upward

Boards do not need twelve metrics. They need three or four, explained plainly, with direction over time.

I suggest: open KEV-listed vulnerabilities on internet-facing systems and time to fix them; time to contain confirmed incidents; AI governance coverage; and one sentence on what the AI got wrong this quarter and what was done about it.

That last item builds trust. A program that reports its own failures is a program leadership can believe.

Measurement pitfalls

A few traps catch almost every team that starts measuring AI in security operations.

Measuring only what the tool reports. Vendors are happy to tell you how many alerts their product handled. That is activity. Measure the outcomes in your own data.

Changing two things at once. If you introduce an AI tool and reorganize the team in the same month, you will not know which change caused the result. Stagger changes when you can, and note them when you cannot.

Ignoring the denominator. Fewer incidents may mean better defense. It may also mean worse detection. Pair every count with a coverage measure, such as the share of critical systems sending logs.

Rewarding speed alone. If analysts are judged only on how fast they close alerts, they will close alerts fast. Pair speed metrics with quality metrics, such as the number of true incidents first suggested for closure.

Treating survey figures as targets. The figures from IBM, Verizon and ISC2 in this paper describe their samples. Your goal is steady improvement against your own baseline, not matching a global average.

Measurement done well is a discipline, not a dashboard. It is how a stretched team proves to itself, and to its leadership, that it is getting stronger.

Part IX: Illustrative scenarios

The three scenarios below are **illustrative and hypothetical**. They do not describe any real organization, client or incident. No figures in them are measured results. They exist to show how the ideas in this paper fit together in practice.

ILLUSTRATIVE · HYPOTHETICAL

Illustrative scenario A (hypothetical): The quiet administrator account

Setting. A regional services company with about 600 staff and a three-person security team. Logs from domain controllers and the VPN are centralized. The team has piloted an AI assistant for log analysis, approved under its governance rules and limited to read-only access.

What happens. The assistant's weekly baseline summary notes that an IT administrator account has been signing in through the VPN at consistent times, but from an address range not seen before, and using built-in administrative tools on a file server where that account has not previously been active. No single alert fired. Every action stayed under existing thresholds.

What the people do. An analyst reviews the summary, checks the raw logs directly and confirms the pattern. She contacts the administrator by phone, using a number from the staff directory rather than one in an email. The administrator has not been working remotely. The team escalates to incident response, follows its plan, and the incident owner decides on containment steps.

Lessons. The assistant did not detect the intrusion. It surfaced a deviation from a baseline that only existed because logging was centralized. The person decided. Verification used an out-of-band channel, an influence-resistant habit. This is the threshold-management problem from Part I, met by baselines and judgment rather than by a single rule.

ILLUSTRATIVE · HYPOTHETICAL

Illustrative scenario B (hypothetical): The patch list nobody trusted

Setting. A mid-size manufacturer with hundreds of open vulnerability findings and a long-standing argument between security and operations about what to patch first.

What happens. The security lead pilots AI-assisted prioritization. The tool combines KEV status, internet exposure and asset importance, and produces one ranked list per system owner, with a short explanation for each item. Two internet-facing remote-access appliances with KEV-listed vulnerabilities rise to the top. One is past its supported lifecycle.

What the people do. System owners review the lists. Operations flags one server that cannot be rebooted during a production run; the owner documents an exception with a compensating control and an expiry date. The appliances are patched within the week. Leadership approves a replacement plan for the end-of-life device.

Lessons. The value was not the score. It was a shared, explained list that owners could accept. The AI ranked; people approved. The team denied the adversary the cheapest route in, the asymmetric move

discussed in Parts I and IV. The team now tracks time to remediate KEV-listed vulnerabilities as a board metric.

ILLUSTRATIVE · HYPOTHETICAL

Illustrative scenario C (hypothetical): The helpful assistant that read the wrong email

Setting. A professional services firm uses an AI assistant to summarize reported phishing emails for its small security team. During the pilot, the assistant also had permission to add sender domains to a block list automatically.

What happens. A reported email contains hidden text crafted to manipulate any AI system that reads it. The assistant's summary describes the message as benign, and the instructions in the hidden text attempt to steer the assistant's output. Because the block-list permission existed, the team realizes that a similar message could have caused the assistant to take an unwanted action.

What the people do. An analyst notices the summary does not match the message's obvious signs of phishing. The team reviews the design under its governance rules, removes the assistant's ability to change the block list, and separates the summarizing function from any function that takes action. They add this case to analyst training and to the scope of their next authorized security test.

Lessons. This is prompt injection, the first risk on the OWASP 2025 list, meeting excessive agency. The fix was architectural, not heroic: reduce the tool's authority, keep a person between output and action, and test the workflow. Over-reliance would have missed it. A trained reviewer caught it.

Across all three scenarios, the pattern is the same. Fundamentals first. AI as an assistant to judgment, not a replacement for it. Clear authority. Measured results.

Recommendations

Here is the plan.

It is not glamorous. It is built around layers, dispersion and redundancy, with no single point that loses the position when it fails. CSIS describes the same logic for communications: dispersed nodes and resilient meshes that keep working "even if multiple nodes go dark" (CSIS, 2025a).

Add capability only when the controls and the evidence are in place. Timelines depend on your team size, your tooling and your risk appetite. The phases below are a sequence first and a schedule second.

The six steps

Phase 0: Govern before you buy. Name one accountable owner for AI in security. Publish an approved-tools list and a plain rule on what data may go into AI services. Use the NIST AI RMF to structure risk decisions. Put role-based access and multifactor authentication on AI models and the data behind them. IBM found that 92% of organizations with an AI-related breach lacked proper AI access controls (IBM and Ponemon Institute, 2026).

Phase 0: Measure a baseline. Write down today's numbers: time to triage an alert, time to identify and contain incidents, time to remediate KEV-listed vulnerabilities, analyst hours spent on reporting. No baseline, no proof. It is that simple.

Phase 1: Pilot with analysts in the loop. Start small and low-risk: alert summaries, indicator enrichment, report drafts, help writing log queries. AI proposes. A person decides. Keep a record of AI outputs and analyst overrides so you can review quality.

Phase 2: Point AI at exposure. Exploitation of vulnerabilities is now the leading way into breaches, so use AI to help prioritize fixes. Combine KEV status, internet exposure and asset importance, and map findings to ATT&CK techniques. System owners keep patch approval. And expect more patches to review, as Verizon does.

Phase 2: Secure the AI itself. Review the APIs, plug-ins and cloud settings around every AI workload. IBM's most common root causes were compromised connected apps, APIs or plug-ins and cloud misconfigurations. Inventory the service accounts and keys your AI tools use, and give each one only the access it needs.

Phase 3: Test it, then widen it. Commission authorized vulnerability assessments and penetration tests that include your AI-enabled workflows, scoped and approved in writing. IBM lists offensive security testing among factors associated with lower breach costs. Extend automation, such as containment steps, only where tests and metrics show it behaves as intended. Keep an easy way for a person to step in.

THE TAKEAWAY

Govern first. Measure second. Point AI at the work that shrinks your exposure. Let evidence, not enthusiasm, decide how much autonomy it gets.

What the battlefield teaches the plan

Each lesson from The modern battlefield maps to a part of the plan. Use this as a check on your own program.

Battlefield lesson	What it teaches	Where it lands in the plan
Persistence beats brilliance	Plan a campaign, protect your people, fund year two	12-month roadmap; staffing and workload metrics; AI for repetitive work only
Cheap mass versus exquisite systems	Answer volume with layers and cheap automated controls	KEV-first patching; MFA; email controls; AI triage at volume, human decision
You are being watched	Know what is visible, reduce it, build your own sensors	Asset and exposure inventory; logging coverage; attack-surface review
Signals can be jammed and spoofed	Verify out of band; keep a second source	Call-back rules; trusted time; cross-checking alerts; model input treated as untrusted
Measure, countermeasure, repeat	Controls decay; schedule the re-test	Quarterly detection review; re-baselining; AI tool reviews; exercises
The rear area is not safe	Your suppliers and platforms are terrain	Dependency map; vendor evidence; offline copies; manual fallbacks
Decentralize decisions, protect the network	Pre-authorize; keep comms that do not depend on what you defend	Written decision rights; out-of-band incident channel; AI as one node, not the spine

And the dead reckoning check, to run each quarter:

1. **Last known position.** Is your baseline current?
2. **Heading.** Are your top priorities written down and still the right ones?
3. **Speed and time.** Do you know your time to detect, contain and fix, and do you trust your time source?
4. **A second source.** For each critical signal, is there an independent check?
5. **A fix.** When did you last re-verify against reality, and correct course?

The first 90 days

The first 90 days are about foundations and a safe, measured pilot. Here is a realistic plan for a mid-size organization. Adjust the pace to your capacity.

Days 1 to 30: Govern and baseline

- Name the accountable owner for AI in security operations.
- Inventory AI tools already in use, including AI features switched on inside existing security and productivity products.
- Publish an approved-tools list and a one-page data rule.
- Confirm multifactor authentication and role-based access on every approved AI tool.
- Record baseline metrics from Part VIII: KEV remediation time, open KEV count on internet-facing systems, logging coverage, time to first decision on alerts, reporting hours.
- Fix the obvious: patch internet-facing systems with KEV-listed vulnerabilities; confirm central logging for authentication, VPN and domain controllers; identify end-of-life technology.

Days 31 to 60: Pilot low-risk use cases

- Choose two pilot use cases. I suggest report drafting and alert-triage summaries.
- Write a use-case card for each: purpose, inputs, outputs, reviewer, prohibited actions, off switch.
- Run in recommend-only mode. No AI-driven closures, no AI-driven changes.
- Log every AI output and every analyst override.
- Hold a weekly fifteen-minute quality review.

Days 61 to 90: Measure, decide, extend to exposure

- Compare pilot metrics against baseline.
- Decide, in writing, whether each pilot continues, changes or stops.
- Begin AI-assisted vulnerability prioritization in recommend-only mode, with system owners approving patch order.
- Run one tabletop exercise in which an AI tool is wrong or unavailable.
- Brief leadership with three or four metrics and one honest note on what the AI got wrong.

The first 12 months

After 90 days, the work becomes a cycle: extend carefully, test, measure, adjust.

Period	Focus	Key activities	Gate to move on
Months 1–3	Govern, baseline, pilot	Owner named; tools inventoried; data rule; baseline metrics; two low-risk pilots; KEV and logging fundamentals	Pilots measured against baseline; written decision on each
Months 4–6	Exposure and log analysis	AI-assisted vulnerability prioritization in production with owner approval; log-analysis assistant with approved data only; enrichment with source discipline	KEV remediation time improving against baseline; query corrections tracked

Period	Focus	Key activities	Gate to move on
Months 7–9	Secure and test the AI	Review APIs, plug-ins and cloud settings around AI workloads; least-privilege service accounts; authorized assessment covering AI-enabled workflows	Test findings remediated or accepted in writing
Months 10–12	Selective automation and early warning	Narrow, reversible automation where evidence supports it; consider deception and early-warning program with legal review; annual governance review	Override and error rates acceptable; off switch tested

A few rules hold across the whole year.

- **Every widening of authority is a decision**, made by the accountable owner, documented, and reversible.
- **Every AI incident triggers a review**, however small.
- **Every quarter, report to leadership** on metrics and on what went wrong.
- **Every year, re-run the maturity self-assessment** below.

Maturity self-assessment checklist

Use this checklist with your security, IT and leadership teams. Answer honestly. A “no” is not a failure. It is a work item.

Score each item: 0 = not in place, 1 = partly in place, 2 = fully in place and evidenced. The levels at the end are my rough guide, not a formal standard.

Governance

☐ An accountable owner for AI in security operations is named and has authority to pause or retire tools.

☐ An approved AI tools list exists and staff know where to find it.

☐ A written rule states what data may and may not go into AI services.

☐ Every AI-enabled tool is inventoried with an owner, data access and service accounts.

☐ Each AI use case has a one-page card with purpose, reviewer, prohibited actions and off switch.

☐ AI risk decisions reference the NIST AI RMF and, for generative tools, NIST AI 600-1.

Fundamentals

☐ Internet-facing systems are inventoried.

☐ KEV-listed vulnerabilities on internet-facing systems are tracked and prioritized.

☐ Phishing-resistant multifactor authentication protects remote access and administrators.

☐ Authentication, VPN, domain controller and critical application logs go to a central store.

☐ Technology beyond its supported lifecycle is identified with a replacement or isolation plan.

☐ Backups of critical systems are tested by actual restoration.

AI in operations

☐ Baseline metrics were recorded before any AI pilot.

☐ AI outputs and analyst overrides are logged.

☐ No alert is closed and no change is made on AI recommendation alone, unless an evidence-based decision allows it for a narrow case.

☐ Vulnerability prioritization uses KEV status, exposure and asset importance, with owner approval.

☐ Threat-intel enrichment cites sources, and attribution claims are treated as low confidence by default.

Securing the AI

☐ APIs, plug-ins and cloud settings around AI workloads have been reviewed.

☐ AI service accounts and keys follow least privilege.

☐ Tools that read untrusted content are separated from tools that take privileged action.

☐ Each AI tool can be switched off quickly by a named person, and that has been tested.

☐ Authorized security testing includes AI-enabled workflows.

☐ Use of unapproved AI services is monitored.

People and resilience

☐ Analysts regularly work cases without AI assistance.

☐ Analysts are trained on prompt injection, over-reliance and data rules.

☐ A tabletop exercise with an AI-failure injection has been run in the past year.

☐ Staff outside security have simple verification procedures for payment and identity requests.

☐ Leadership receives quarterly metrics including what the AI got wrong.

Reading your score

There are 28 items, for a maximum of 56 points.

- **0–19: Foundational.** Focus on fundamentals and governance before expanding AI use.
- **20–39: Developing.** Pilots are reasonable. Prioritize exposure reduction and securing the AI.
- **40–56: Established.** Consider selective automation and early-warning programs, with continued testing.

Whatever your score, the next step is the same: pick the three lowest items that matter most to your risk, assign owners and dates, and re-score in a quarter.

Conclusion

The research agrees on direction.

Attackers are faster. Known vulnerabilities stay open too long. Security teams are short of people and skills. And organizations that use AI and automation extensively report lower costs and faster response.

It agrees on risk, too. AI systems and unsanctioned AI use create new exposure when access control and governance lag behind adoption.

The irregular warfare lens explains why this matters so much. Your opponents, whatever their motive, behave the way doctrine describes. They work indirectly. They avoid attribution. They impose costs asymmetrically. They stay under your thresholds and exploit the seams between your teams, your suppliers and your decisions. AI gives them more speed, scale and intensity.

The public analysis of the modern battlefield points the same way. Persistence. Observation. Cheap mass. Jammed and spoofed signals. Supply lines that are never safe (CSIS, 2025a; CSIS, 2025b). Your organization is not at the front. But it works in the same electromagnetic and informational weather.

You will not win that contest by matching them effort for effort.

When the signal fails, you navigate by dead reckoning. You know your last position. You know your heading. You keep a second source. You correct when you can.

You win it the way resilient institutions have always won it. Protect what matters most. Deny the easy routes. Organize so that decisions are clear and fast. Learn faster than the other side.

So for a mid-size organization, the winning move is neither to avoid AI nor to hand it the keys. Adopt it in stages. Govern, measure, pilot with people in the loop, shrink your exposure, secure the AI itself, and widen automation as the evidence supports it. Done this way, AI stops being another tool to manage and becomes one integrated part of your security program.

That is what practical, integrated security looks like.

Infinitum Corporation helps organizations take these steps in a way that fits their size, systems and risk appetite. I make no promise about the outcome any organization will achieve, and every figure in this paper belongs to the source cited.

Don’t take my word for it. Read the sources, then decide.

This is the first paper in a series. The next two are *Trenches and Drones: Network Defense for a War of Attrition* and *Break It Before They Do: Authorized Hacking for Organizations Under Siege*. Both build on the foundation laid here.

Let's stay engaged. Call us at (202) 980-9300, or visit https://www.theinfinitumcorporation.com. Dialogue is key.

About Infinitum Corporation

Infinitum Corporation is a technology and security consulting firm headquartered in Washington, DC, delivering practical, integrated security solutions.

Our work spans cybersecurity consulting, network architecture and defense, vulnerability assessment and ethical hacking, IT security, AI for security operations, strategic risk assessment and emerging defense technology advisory. Our ethical hacking work is always authorized, scoped and approved in writing, and performed for defense.

William Aaron Turner is Founder and President.

Website	https://www.theinfinitumcorporation.com
Phone	(202) 980-9300
Telegram	https://t.me/infinitumcorporation
X	https://x.com/infinitumceo
Address	1629 K Street NW, Suite 300, Washington, DC

References

1. CISA (Cybersecurity and Infrastructure Security Agency). *Known Exploited Vulnerabilities Catalog.* https://www.cisa.gov/known-exploited-vulnerabilities-catalog
2. CISA, NSA, FBI, et al. (2024, February 7; updated March 7, 2024). *Joint Cybersecurity Advisory AA24-038A* (advisory title abbreviated here to keep this paper party-neutral; see the URL for the full title). https://www.cisa.gov/news-events/cybersecurity-advisories/aa24-038a
3. Congressional Research Service (Theohary, C. A.). (2024, November 29). *Defense Primer: What Is Irregular Warfare?* (IF12565). https://www.congress.gov/crs-product/IF12565
4. CSIS (Sheppard, L. R., Friend, A. H., Shah, H., Akca, A., Hicks, K. H., & Federici, J.). (2019, July 8). *By Other Means Part I: Campaigning in the Gray Zone.* Center for Strategic and International Studies. https://www.csis.org/analysis/other-means-part-i-campaigning-gray-zone
5. European Commission and High Representative of the Union for Foreign Affairs and Security Policy. (2016, April 6). *Joint Framework on countering hybrid threats: a European Union response* (JOIN(2016) 18 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52016JC0018
6. Hybrid CoE (European Centre of Excellence for Countering Hybrid Threats). *Hybrid threats as a concept.* https://www.hybridcoe.fi/hybrid-threats-as-a-phenomenon/
7. IBM and Ponemon Institute. (2026). *Cost of a Data Breach Report 2026.* https://assets.ibm.com/adobe/assets/urn:aaid:aem:21111142-1251-4369-86fb-57b82f5bb108/original/as/Cost%20of%20a%20Data%20Breach%20Report%202026.pdf
8. IBM Newsroom. (2026, July 29). *IBM Study: One in Four Malicious Breaches are AI-Enabled, Costing Companies $6 Million on Average.* https://newsroom.ibm.com/2026-07-29-ibm-study-one-in-four-malicious-breaches-are-ai-enabled,-costing-companies-6-million-on-average
9. ISC2. (2024, October 31). *2024 ISC2 Cybersecurity Workforce Study.* https://www.isc2.org/Insights/2024/10/ISC2-2024-Cybersecurity-Workforce-Study
10. MITRE. *ATLAS: Adversarial Threat Landscape for AI Systems.* https://atlas.mitre.org/; content release 2026.09 (2026, September 14), https://github.com/mitre-atlas/atlas-data/blob/main/CHANGELOG.md
11. MITRE. *ATT&CK.* https://attack.mitre.org/
12. MITRE. *Engage: An Adversary Engagement Framework.* https://engage.mitre.org/
13. NATO. (2026, January 29, last updated). *Countering hybrid threats.* https://www.nato.int/cps/en/natohq/topics_156338.htm
14. NIST. (2023). *AI Risk Management Framework (AI RMF 1.0)*, including the Generative AI Profile, NIST AI 600-1 (2024, July 26). https://www.nist.gov/itl/ai-risk-management-framework
15. NIST. (2023). *AI RMF Core* (Section 5 of AI RMF 1.0). Trustworthy and Responsible AI Resource Center. https://airc.nist.gov/airmf-resources/airmf/5-sec-core/
16. NIST. (2024, February 26). *NIST Releases Version 2.0 of Landmark Cybersecurity Framework.* https://www.nist.gov/news-events/news/2024/02/nist-releases-version-20-landmark-cybersecurity-framework; framework home: https://www.nist.gov/cyberframework
17. NIST (Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., & Hamin, M.). (2025, March). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations* (NIST AI 100-2 E2025). https://csrc.nist.gov/pubs/ai/100/2/e2025/final
18. NSA Artificial Intelligence Security Center, CISA, FBI, et al. (2024, April 15). *Deploying AI Systems Securely.* https://www.cisa.gov/news-events/alerts/2024/04/15/joint-guidance-deploying-ai-systems-securely

19. OWASP Gen AI Security Project. (2025). *Top 10 for LLM Applications 2025.* https://genai.owasp.org/llm-top-10/

20. RAND (Morris, L. J., Mazarr, M. J., Hornung, J. W., Pezard, S., Binnendijk, A., & Kepe, M.). (2019, June 27). *Gaining Competitive Advantage in the Gray Zone: Response Options for Coercive Aggression Below the Threshold of Major War* (RR-2942-OSD). https://www.rand.org/pubs/research_reports/RR2942.html

21. U.S. Army Center for Army Lessons Learned. (2025, July). *Understanding Irregular Warfare* (No. 25-897), quoting JP 1 Vol 1, *Joint Warfighting* (2023, August 27). https://www.army.mil/article/286976/understanding_irregular_warfare

22. Verizon. (2026). *2026 Data Breach Investigations Report.* https://www.verizon.com/business/resources/Te26/reports/2026-dbir-data-breach-investigations-report.pdf

23. Verizon News. (2026, May 19). *Vulnerability exploitation top breach entry point, 2026 DBIR finds.* https://www.verizon.com/about/news/breach-industry-wide-dbir-finds

24. CSIS (Pusztaszeri, A., & Harding, E.). (2025, September 16). *Technological Evolution on the Battlefield*, in the CSIS report *War and the Modern Battlefield* (subtitle omitted to keep this paper party-neutral). Center for Strategic and International Studies. https://www.csis.org/analysis/chapter-9-technological-evolution-battlefield (cited as CSIS, 2025a)

25. CSIS (Bingen, K. A.). (2025, September 16). *Extending the Battlespace to Space*, in the CSIS report *War and the Modern Battlefield* (subtitle omitted to keep this paper party-neutral). Center for Strategic and International Studies. https://www.csis.org/analysis/chapter-8-extending-battlespace-space (cited as CSIS, 2025b)

26. Council of the European Union (General Secretariat). (2025, May 21). *Call for common actions in response to Global Satellite Navigation Systems (GNSS) jamming and spoofing threats* (note 9188/25, from three member states). https://data.consilium.europa.eu/doc/document/ST-9188-2025-INIT/en/pdf

27. Traficom (Finnish Transport and Communications Agency). (2025, June 13). *Finland and Estonia warn vessels in the Gulf of Finland of an increase in GNSS disruptions.* https://traficom.fi/en/news/finland-and-estonia-warn-vessels-gulf-finland-increase-gnss-disruptions

28. EASA and IATA. (2025, June 18). *EASA and IATA outline comprehensive plan to mitigate GNSS interference risks.* https://www.easa.europa.eu/en/newsroom-and-events/press-releases/easa-and-iata-outline-comprehensive-plan-mitigate-gnss

29. ENISA. (2025, October 1). *EU consistently targeted by diverse yet convergent threat groups* (ENISA Threat Landscape 2025 press release). https://www.enisa.europa.eu/news/etl-2025-eu-consistently-targeted-by-diverse-yet-convergent-threat-groups

30. NSA AI Security Center, CISA, FBI, et al. (2025, May 22). *AI Data Security: Best Practices for Securing Data Used to Train & Operate AI Systems* (Cybersecurity Information Sheet). https://www.ic3.gov/CSA/2025/250522.pdf

Note on in-text tags. Short tags such as (Verizon, 2026), (IBM and Ponemon Institute, 2026), (ISC2, 2024), (CRS, 2024), (CSIS, 2019), (RAND, 2019), (NATO, 2026), (Hybrid CoE), (CISA et al., 2024), (NIST, AI RMF), (NIST, AI RMF Core), (NIST, 2024), (NIST, 2025), (OWASP, 2025), (MITRE ATLAS, 2026), (MITRE Engage), (NSA AISC et al., 2024), (U.S. Army CALL, 2025), (CSIS, 2025a), (CSIS, 2025b), (Council of the EU, 2025), (Traficom, 2025), (EASA, 2025), (ENISA, 2025) and (CISA et al., 2025) refer to the entries above. Forecasts, opinions and illustrations attributed to a source are that source's own views. All sources were opened and checked on October 25, 2026. The tag (CISA et al., 2024) is advisory AA24-038A (entry 2); (CISA et al., 2025) is the AI data security sheet (entry 30).

Glossary

AI agent. An AI system that can plan and take actions, such as running queries or changing settings, rather than only producing text.

AI RMF. The NIST Artificial Intelligence Risk Management Framework, a voluntary framework organized around four functions: govern, map, measure and manage.

Asymmetric cost imposition. A situation in which one side spends little to force the other side to spend a lot. Joint doctrine describes irregular warfare as seeking to increase risk and costs to adversaries.

ATLAS. MITRE's knowledge base of adversary tactics and techniques against AI-enabled systems.

ATT&CK. MITRE's knowledge base of adversary tactics and techniques based on real-world observations, used to map detections and find coverage gaps.

Attribution. Determining who is responsible for an activity. Gray-zone and hybrid activity is often designed to make attribution difficult.

Authorized testing (ethical hacking). Security testing of systems performed only with explicit, written permission from the system owner, within an agreed scope and rules of engagement, for the purpose of defense.

Automation bias. The tendency of people to over-trust automated recommendations, especially after the system has been right many times.

Baseline. A documented picture of normal activity or performance, used to spot deviations and to measure improvement.

CSF 2.0. The NIST Cybersecurity Framework version 2.0, organized around six functions: Govern, Identify, Protect, Detect, Respond and Recover.

Data and model poisoning. Manipulation of the data used to train, fine-tune or inform an AI system so that it behaves incorrectly.

Deception (defensive). Placing decoy assets, credentials or data inside your own environment so that intruders who interact with them reveal themselves.

Excessive agency. Granting an AI system more ability to act than its task requires.

Gray zone. The contested space between routine statecraft and open warfare, in which actors pursue objectives while staying below thresholds that would trigger a conventional military response.

Hybrid threats. Coordinated combinations of military and non-military, covert and overt means, such as disinformation, cyber attacks and economic pressure, used to achieve objectives below the threshold of declared war.

Irregular warfare (IW). In current U.S. joint doctrine, a form of warfare where states and non-state actors campaign to assure or coerce states or other groups through indirect, non-attributable, or asymmetric activities.

KEV. CISA's Known Exploited Vulnerabilities catalog, a list of vulnerabilities known to have been exploited in the wild.

Living off the land (LOTL). The use of tools already present on a system, rather than introduced malware, to carry out malicious activity, making it harder to distinguish from legitimate administration.

Penetration test. An authorized, scoped exercise in which testers attempt to identify and safely demonstrate exploitable weaknesses, under written permission, so they can be fixed.

Phishing-resistant MFA. Multifactor authentication methods designed to resist credential theft through phishing.

Pre-positioning. Gaining and maintaining access to a network in advance, for possible later use, rather than for immediate effect.

Prompt injection. Content supplied to an AI system, directly or hidden inside processed material, that alters its behavior in ways its operators did not intend.

Proxy. An intermediary actor or piece of infrastructure used to carry out or conceal activity on behalf of another party.

Resilience. The ability of an organization to absorb disruption, keep essential functions running and recover.

Shadow AI. The use of AI tools by staff without organizational approval or oversight.

SOC. Security operations center: the team and capability that monitors, detects, investigates and responds to security events.

Threshold management. Deliberately keeping activity below the levels that would trigger detection or response.

Vulnerability assessment. A systematic review of systems to identify, rank and report security weaknesses, without necessarily attempting to exploit them.