HILLSTRONG GROUP SECURITY

HILLSTRONG · THOUGHT LEADERSHIP

OT Security and the Age of AI

From Automation to Accountability

Roger Hill | CEO, Hillstrong Group Security, Inc.

Contents

CHAPTER ONE	The Convergence Imperative: Why AI Collapses Security Silos
CHAPTER TWO	OT Security Meets AI: Blind Spots in the Plant Floor
CHAPTER THREE	When Cyber-Physical Systems Collide with Privacy and IP: Preparing Manufacturing for AI's Next Chapter
CHAPTER FOUR	Ethics & Accountability in Industrial AI: Why Trust is Your Hardest Asset to Build
CHAPTER FIVE	Building Resilient, Explainable AI Systems: The Hard Truth About Trustworthy Autonomy
	About the Author
	About Hillstrong Group Security, Inc.

HILLSTRONG GROUP SECURITY

The Convergence Imperative: Why AI Collapses Security Silos

AI is not an OT feature, it is a governance test that merges security, privacy, and accountability.

Not another feature

Most of the AI messaging aimed at industry is framed as incremental progress. Vendors pitch smarter anomaly detection, predictive maintenance, or faster inspection. The sales language makes it sound like a routine software upgrade.

That framing is misleading. AI is not just another feature. It moves data in ways plants have never had to govern, and it erases the boundaries executives rely on to divide responsibility. Information that used to stay within OT networks is now shared with IT systems, pushed into cloud pipelines, and ingested by models that may retain it. Logs once used for troubleshooting can be summarized by large language models (LLMs) that inadvertently surface personal details. Engineering data uploaded for support may end up outside your control.

AI collapses the walls between OT security, IT security, privacy, and corporate governance. The question for boards is whether the organization is ready for that collapse.

How boundaries erode

For decades, manufacturing risk was managed through specialization. OT security focused on safety and uptime. IT security defended infrastructure and corporate information. HR and legal handled privacy. These groups rarely collided.

AI changes that equation. Even modest pilots create flows that jump lanes:

- Predictive systems collect operational data and store it in external environments.
- OT alerts arrive in IT-managed security operations centers, where staff lack industrial context.
- Maintenance notes, often containing names or supervisor comments, are ingested by LLMs.

It doesn't take widespread deployment to create exposure. The simple act of experimenting with AI forces executives to govern across functions.

What convergence looks like in a plant

Operational data as performance data

Smart torque tools, barcode stations, and badge-gated checkpoints already log who did each step, how long it took, and in what sequence. When AI models use this data to optimize cycle time, the result isn't just efficiency, it becomes a profile of individual or team performance. That is privacy-sensitive information whether executives label it that way or not.

Maintenance notes as personal records

Enterprise Asset Management (EAM) systems and root cause analysis (RCA) databases contain years of incident reports, work orders, and comments. Training an LLM on this corpus can be useful for spotting trends. It can also resurface employee names, shift identifiers, and even blame language. What started as a maintenance project quickly becomes a labor and privacy issue.

Configuration data as intellectual property

Under pressure to restore uptime, engineers may upload Programmable Logic Controller (PLC) backups, recipes, or network diagrams into an AI-powered support assistant. Those files encode tolerances, sequencing, and architecture, the design DNA of the operation. If the vendor retains them for "model improvement," a company's crown jewels have effectively been transferred.

Each of these examples is operationally realistic. Each one shows how quickly AI pilots can cross into privacy, security, or intellectual property territory.

Misjudgments I hear from leadership

In executive discussions, four themes repeat:

- "This belongs to OT." It doesn't. AI crosses IT and privacy boundaries on day one.
- "It's only a pilot." Pilots normalize behaviors, where data is sent, what is retained, that are hard to undo later.
- "Privacy doesn't apply here." It does once telemetry or text corpora can be linked to individuals.
- "The vendor will cover us." Vendors design contracts to disclaim liability. Boards inherit the risk.

These are not trivial misunderstandings. They represent a blind spot that will eventually expose the organization if governance is not reset.

How this plays out in real scenarios

Imagine three common situations:

- A cycle-time optimization tool uses smart tool telemetry. Months later, workers ask why the company is ranking them without disclosure or consent.
- A predictive maintenance pilot trains on years of shift notes. The AI model produces summaries that include supervisor critiques of specific employees. Legal and HR are drawn in after the fact.
- A troubleshooting request involves uploading PLC backups to a vendor. The files are retained, and later the company realizes proprietary tolerances have left its control.

None of these scenarios are exotic. They are natural consequences of how AI consumes the data plants already produce.

The governance gap

What makes AI different is not the features, it's the accountability vacuum. Vendors market efficiency but decline liability. IT and OT functions push responsibility back and forth. Privacy is often consulted too late.

Without governance in place, early projects set their own rules by convenience. Data leaves plants without review. Logs are retained without oversight. Sensitive files are uploaded under pressure. By the time boards intervene, practices are entrenched.

This is the *real* risk: AI doesn't just collapse silos of data. It exposes silos of accountability.

What boards can do now

The right time to act is before adoption accelerates. Boards should:

- Create a governance policy. Define what operational, personal, and configuration data can be used in AI projects and what protections are mandatory.
- Run impact reviews. Assess not only what data is collected but also what inferences are generated.
- Set strict vendor terms. Require contracts that prohibit retention or reuse of data unless explicitly authorized.
- Name a responsible executive. Decide whether AI governance belongs to the CIO, CISO, or COO. Vagueness guarantees gaps.

What this means for directors

AI will not remain confined to pilots for long. The pressure to show efficiency gains will push adoption forward. Once it does, the old divisions between OT, IT, privacy, and governance won't hold.

The boards that prepare now will adopt AI on their own terms, with clear policies, contracts, and accountability. The boards that delay will discover that AI adoption has already rewritten the rules without them.

The test is straightforward: do you treat AI as another plant feature, or do you see it as the governance stress test it really is?

OT Security Meets AI: Blind Spots in the Plant Floor

AI cannot protect what it cannot see, and in plants, the most important signals are often invisible

The flashlight problem

AI in OT security is being marketed as a breakthrough in visibility. Dashboards fill with asset inventories, anomaly scores, and traffic maps that suggest comprehensive coverage. For executives responsible for resilience, it looks like a long-awaited solution.

But AI is not a floodlight that reveals every corner of a plant. It is a flashlight. It illuminates what you point it at and leaves the rest in shadow. That distinction matters. The risk is not that coverage is partial, executives can manage that. The risk is when partial coverage is mistaken for total coverage, and leaders make decisions on the belief that blind spots no longer exist.

Where the light doesn't reach

The blind spots in OT environments are structural. They are not a failure of AI; they are a reality of how industrial systems were designed.

Consider the control cabinet. Many still use unmanaged switches that do not support port mirroring. The conversations between a Programmable Logic Controller (PLC), a Human–Machine Interface (HMI), and a drive never leave the box. If sensors are positioned at the network core, they will never capture that traffic. The AI model built on those feeds has nothing to analyze.

Or take the engineering workstation. A controls engineer connects directly to a PLC for a quick update. That session is point-to-point. It bypasses taps, spans, and anomaly detection entirely. If malware moves across that session or logic is changed, there will be no record.

Logs pose a similar problem. Systems such as SCADA servers, historians, and MES platforms generate records that can help detect misuse. Yet many plants do not centralize or retain these logs consistently. AI log analysis can only work with what is collected. When the data is absent, the analysis is silent.

Even SIEM, which in IT aggregates security events across the enterprise, faces limits in OT. Embedded devices like PLCs, DCS controllers, and motion systems were never designed to generate SysLog or similar event streams. They simply do not speak. The only practical alternative is to monitor the PC-based systems that interact with them, engineering workstations, SCADA servers, operator HMIs, historians. Those endpoints act as proxies. It is valuable, but it is indirect and incomplete.

How the message changes on its way up

At the engineering level, these blind spots are well understood. Network architects know which cabinets cannot be mirrored. Controls teams know which laptops connect directly. Security specialists know which logs exist and which don't.

When security teams evaluate anomaly detection platforms, they work within those constraints. They collect what they can, deploy endpoint tools where possible, and feed available logs into SIEM. The dashboards look promising. Asset inventories expand. Baselines stabilize.

But as the information moves upward, the nuance erodes. P&L owners, the people approving spend, hear about improved visibility and AI-driven insights. The technical caveats often do not make it into the briefing. By the time results are presented to a board, what survives is the headline: "We've deployed AI to increase visibility and resilience."

This is not intentional distortion. It is compression. Complex technical realities are flattened into business language. The blind spots are still there, but they are no longer part of the conversation.

False assurance as the real risk

Every plant has blind spots. That is expected. The real concern is when those blind spots disappear from the picture leaders use to make decisions.

Partial visibility can be managed. Executives can set compensating controls, govern accordingly, and disclose residual risk. False visibility is more dangerous. It creates misplaced confidence. Budgets shift elsewhere. Controls are relaxed. Risk statements to regulators and insurers overstate resilience. Then, when an incident occurs in the unseen corners, the gap between perception and reality becomes painfully clear.

This is the governance failure to guard against. Not that AI misses things, that is a given, but that its limitations are not consistently surfaced in the way assurance is communicated.

A scenario to test the claim

Picture a manufacturer that deploys AI-enabled anomaly detection across several plants. Dashboards fill with new asset inventories, baseline maps, and anomaly scores. The business case cites expanded visibility, and the rollout is approved for more sites.

Months later, a PLC fails with a fault condition following a rushed programming change. The engineer connected directly from a laptop, bypassing all monitoring systems. No logs were captured, and the AI system raised no alert.

On the plant floor, this is no surprise. Engineers knew direct connections are invisible. Security teams may have flagged it in risk registers. But those caveats never made it into the business case that justified the spend.

What reached executives was "visibility improved." What reached the board was "resilience enhanced." Both statements were technically accurate. Neither told the whole story.

What leaders should actually expect

AI promises improved anomaly detection. Vendors claim it will accelerate asset discovery, identify out-of-pattern flows, and summarize logs faster than people alone. But these promises cannot rewrite the fundamentals: you cannot analyze data that is never collected, and in OT environments, some data will simply never be collected.

For P&L owners, this means treating AI promises with scrutiny. Ask for evidence of what the platform actually ingests. Require visibility assurance as part of every funding proposal. If a percentage of activity remains unobserved, that percentage should be disclosed.

For boards, the role is not to debate network protocols. It is to test the credibility of assurance. When management reports improved visibility, are they also disclosing what remains unseen? Oversight is about ensuring executives are not compressing nuance into slogans.

A systematic approach to reducing blind spots

The smart path is not to reject AI anomaly detection. It is to apply rigorous system architecture principles to minimize blind spots from the start.

Begin with a comprehensive design analysis of your OT environment. Apply systematic engineering requirements to network architecture, ensuring monitoring capabilities are built into the infrastructure rather than added as an afterthought. Configure systems with visibility as a core requirement, not just operational efficiency. Properly tune detection mechanisms based on criticality assessments of each system component.

Then create a visibility governance framework. Document what is captured and what remains outside view. Quantify coverage with metrics that matter to both engineers and executives. Where monitoring gaps remain despite architectural improvements, implement robust compensating controls: cryptographically signed logic changes, secure jump hosts for programming sessions, and formal attestation procedures. Ensure this comprehensive view is presented whenever visibility is reported up the chain.

AI in OT security is a flashlight, not a floodlight. It is useful. It reveals things that were hidden. But it does not change the architecture of plants, and it does not eliminate blind spots.

The responsibility of leadership is not to expect perfection from the tool but to demand honesty in how its limits are reported. Engineers know where the light doesn't reach. Security teams can articulate it. P&L owners must insist it is disclosed. Boards must ensure the assurances they receive reflect the full picture.

The next time a dashboard makes the plant look completely visible, pause and ask: are we governing to what is actually seen, or to what the dashboard makes us believe?

HILLSTRONG GROUP SECURITY

OT SECURITY AND THE AGE OF AI · CHAPTER THREE

When Cyber-Physical Systems Collide with Privacy and IP: Preparing Manufacturing for AI's Next Chapter

AI and LLMs are beginning to blur operational, privacy, and intellectual property boundaries: boards must set the guardrails now

A change that will catch leaders off guard

Factories were designed to optimize throughput and protect physical safety. The information they generated (sensor readings, control logic, maintenance logs) was treated as technical. It mattered to operators and engineers, but it never drew the attention of regulators, lawyers, or auditors.

That clean division is starting to disappear. As vendors fold AI features into their products and researchers test large language models (LLMs) on operational data, manufacturing executives face a new reality: information once considered harmless can now be classified as personal, regulated, or strategically sensitive.

This shift is not happening at scale yet. But it is emerging fast enough that boards who wait until adoption is widespread will find themselves reacting under duress rather than shaping terms of engagement.

When plant data stops being "just plant data"

One of the quiet surprises of early AI adoption is how ordinary operational data changes character once models begin to process it.

Take cameras. Vision systems installed for quality inspection do more than check product tolerances; they capture images of workers in the frame. Those images may qualify as biometric data.

Consider logs. An LLM trained to summarize maintenance reports doesn't just highlight equipment conditions; it can also surface names, comments, and shift patterns. What was once a troubleshooting record suddenly looks like a labor file.

Even optimization algorithms can push boundaries. A scheduling tool may begin surfacing insights about team performance. Regulators could easily interpret that as a form of surveillance.

None of these examples are science fiction. They are plausible consequences of how AI changes the nature of data. The lesson for executives is simple: once AI enters the loop, operational information can no longer be assumed safe from privacy obligations.

The overlooked crown jewels

Privacy is only part of the story. Intellectual property (the formulas, tolerances, and configurations that define how plants actually run) may be even more exposed.

Many AI-driven services ask for access to detailed technical information. That can mean:

- recipes and formulations developed over decades
- PLC configuration files that encode sequencing and tolerances
- OT network diagrams that show exactly how critical assets are connected

If this material is uploaded into a vendor's AI system, the company has lost control over some of its most valuable assets. Vendors often retain the right to keep data for "model improvement." Once that happens, your competitive advantage may be sitting in someone else's pipeline.

It's not hard to imagine how this unfolds. An engineer, under pressure to get a line back up, sends configuration files to an AI assistant for faster troubleshooting. The files help resolve the problem, but they also reveal how that process is tuned. Unless clear contractual protections are in place, that information may not stay contained.

For boards, this is not a technicality. It is a direct threat to shareholder value.

Data doesn't always stay where you think

Almost every AI service in the industrial space is cloud-first. That means the data collected, whether logs, images, or configurations, often leaves the plant. Even in pilots, it may travel across jurisdictions, where different privacy, labor, or export laws apply.

Few manufacturers today have a clear line of sight on these flows. Questions that rarely get asked include:

- Where is the data physically stored?
- Which subcontractors are processing it?
- What legal regimes govern access once it leaves the country?

The answers matter. A misstep here doesn't just create a compliance problem, it can generate headlines, lawsuits, and a loss of trust with regulators who will expect you to have known better.

The problem isn't just what AI collects, it's what it infers

LLMs are particularly tricky because they don't just store data; they generate new insights. Those insights can reveal more than the underlying records ever did.

Maintenance logs that once documented machine failures might, when analyzed, highlight absenteeism patterns or flag a labor dispute. A scheduling optimizer could suggest vulnerabilities in a supplier's workforce. Quality models might tie defect rates to certain crews, effectively turning process monitoring into performance monitoring.

Executives should assume regulators will view inferences as carrying the same obligations as raw data. A profile created by a model can be just as sensitive as the identifiers it was built from.

Governance is already behind the curve

Most manufacturers do not yet have a board-level AI policy. That is understandable, adoption is limited, and much of what exists is experimental. The problem is that pilots are moving forward anyway. Without governance, early projects establish habits by default: sending logs to vendors, uploading configurations, streaming video feeds without restrictions.

Once those practices normalize, they are hard to undo. Culture and convenience set in. At that point, governance is playing catch-up.

The deeper issue is fragmentation. Vendors disclaim liability. IT leaders see it as an OT concern. OT leaders dismiss it as "just data." Privacy and legal are consulted only when a regulator calls. This lack of ownership is itself a risk. AI doesn't respect organizational silos, and neither will regulators when they assess accountability.

How it could play out in practice

To make this real, consider a few plausible scenarios:

- Defect detection pilots. A food processor tests an AI vision system to flag flawed products. The system also records workers at their stations. If those images are retained, they may be treated as biometric data.
- Shift log summaries. A manufacturer experiments with LLMs to speed up maintenance reporting. The model not only highlights equipment issues but also surfaces names and notes about individual employees. A tool built for efficiency now carries privacy implications.
- Cloud-based anomaly detection. A plant streams OT traffic to a vendor's AI platform. Months later, it becomes clear the data was routed offshore through subcontractors. That discovery creates compliance questions the board wasn't prepared to answer.
- AI-driven support tools. Engineers upload PLC configuration files into a vendor's troubleshooting assistant. Those files encode process IP central to the company's competitive advantage. If they are stored, the vendor has information competitors would pay dearly to see.

These examples aren't predictions, they're cautionary illustrations. They show how quickly operational experiments can bleed into privacy, compliance, and IP exposure.

What boards should do now

The point is not to stop experimentation with AI. The point is to shape it with guardrails. Boards can take several steps immediately:

- Issue an AI governance policy. Even a short one makes clear what data can and cannot be used, how IP must be protected, and who approves exceptions.
- Run impact assessments. Treat AI pilots the same way you would a new system handling customer data. Document risks before deployment.
- Set vendor terms. Contracts should prohibit retention or reuse of your operational data and IP. Do not accept vague promises of "best practices."

- Use frameworks as scaffolding. The NIST AI Risk Management Framework and the EU AI Act offer useful benchmarks. They give boards a way to show regulators that standards were considered early.
- Name a single accountable executive. Decide now who owns AI governance, CIO, CISO, or COO. Shared accountability often means no accountability.

What this means at the board table

The key issue isn't whether AI is pervasive in your plants today. It isn't. The issue is whether you are setting standards before adoption accelerates.

Operational data once treated as harmless is on a trajectory to become regulated or strategically sensitive. AI features, LLMs especially, are the catalyst. They collapse the boundaries between operational security, information security, privacy, and intellectual property protection.

Boards that move early can adopt AI responsibly and defend their competitive position. Boards that wait will face difficult questions about why sensitive data was allowed into vendor pipelines without safeguards.

The next part of this series will focus on accountability. If AI introduces new categories of risk, who owns them? Vendors rarely accept liability. Leaving responsibility undefined is an option no board can afford.

Ethics & Accountability in Industrial AI: Why Trust is Your Hardest Asset to Build

Aligning with NIST AI RMF to Build Trust

Manufacturing lives and dies on trust. Just ask any plant manager. The night supervisor counts on valves opening the moment they're needed. Engineers bet their reputations on PLC code standing firm under pressure spikes. Safety teams stake lives on emergency systems triggering the instant parameters exceed thresholds. When that trust fractures, everything unravels, production rates nosedive, error counts climb, and risk exposure multiplies. Now AI has entered the equation, demanding we extend trust beyond physical components into the abstract realm of algorithms that make high-stakes predictions where errors directly translate to seven-figure losses or, worse, accident reports.

Let's be brutally honest: you can drop the most sophisticated AI platform into your facility tomorrow morning, but if operators, engineers, and regulators don't trust it down to their bones, that platform becomes glorified shelfware. Your multi-million dollar investment returns nothing but gathering dust. And trust isn't manufactured through slick marketing decks, it's earned through transparent ethics and clear lines of accountability visibly woven through every system interaction. No shortcuts exist.

The Industrial AI Trust Gap

AI is no longer at the edge of operations. It is being embedded directly into quality control, predictive maintenance, scheduling, and even OT cybersecurity detection systems. Each of these touches revenue, compliance, and safety. Boards are betting that AI will unlock efficiency and insight, yet few have seriously confronted the ethical and governance questions that come with it.

The emerging NIST AI Risk Management Framework (AI RMF) is not just another government checklist. It represents a shift in how trust in AI is defined. Fairness, transparency, robustness, and accountability are not abstract ideals. In a manufacturing context, they directly shape whether customers, regulators, and employees believe your AI can be trusted in critical operations.

The real challenge is cultural. Many executives see ethics as a compliance issue, an obligation to meet the bare minimum. In reality, it is a strategic differentiator. The firms that can demonstrate ethical stewardship of AI will win customer confidence, regulator goodwill, and operator adoption. Those that cannot will face skepticism, stalled deployments, and hidden risk.

Breaking Down the NIST AI RMF for Industry

The NIST AI RMF defines characteristics of trustworthy AI. Applied to industrial settings, they take on very practical meaning:

- Validity and Reliability
- Transparency and Explainability
- Safety and Robustness
- Fairness and Bias Mitigation
- Accountability and Governance

Critical Moments Where Ethics Can Falter: Hypothetical Examples

- Predictive Maintenance Without Explainability: Consider a hypothetical scenario where a manufacturer deploys AI to forecast motor failures. Operators receive alerts stating: "replace this motor in 48 hours" without supporting rationale. Without understanding why, operators might reasonably dismiss these warnings until an actual failure occurs, potentially resulting in significant financial losses. The issue here wouldn't be model accuracy but rather its opacity.
- Data Privacy in Quality Control: Imagine a food processing company implementing AI-powered cameras to detect product defects. While focusing on quality improvements, the system might inadvertently capture and store personally identifiable worker behaviors. What begins as a process optimization initiative could raise serious labor privacy concerns if proper boundaries aren't established.
- Cybersecurity Blind Spots: Picture an OT security system marketed with AI-driven threat detection capabilities. In this scenario, certain network segments like unmanaged switches remain outside monitoring scope. If vendors overpromise with claims like "AI will automatically adapt" rather than transparently acknowledging limitations, trust erodes when these blind spots inevitably become exposed.

These are not technical failures alone. They are ethical and governance failures.

Common Misconceptions About AI Ethics

Executives often fall into these predictable traps:

- "Ethics is just for compliance." In reality, ethics goes far beyond checking boxes, it builds the confidence necessary for successful adoption.
- "Transparency slows us down." The truth is exactly opposite. Without explainability, operators will reject AI systems, bringing adoption to a complete standstill.
- "The vendor owns ethics." Vendors merely sell models. You own all the risk when those models influence your operations. Ethics must be embedded in your governance, not outsourced to third parties.
- "We'll fix ethical issues after deployment." Trust must be designed from the beginning. Once operators or regulators lose faith in your systems, rebuilding that trust becomes ten times harder.

A Better Way to Frame Ethics in Industrial AI

To make ethics actionable in industry, executives should focus on three pivots:

- From Compliance to Confidence: Treat ethical alignment as a differentiator, not a burden. Customers and regulators increasingly favor firms that demonstrate trustworthy AI. This confidence directly accelerates market advantage.
- From Black Box to Glass Box: Insist on explainability. Engineers must be able to interrogate why the model flagged a failure. This isn't optional, it's the difference between adoption and rejection.
- From Abstract Principles to Operational Discipline: Ethical frameworks must tie directly to real governance practices: retraining schedules, override procedures, and data retention policies. Ethics lives in operational details, not in policy documents.

Practical Actions for Executives

- Document Model Decisions. Require that every AI recommendation is logged with inputs, rationale, and outcome. This creates auditability for regulators and learning for operators.
- Enforce Explainability. Reject models that cannot provide interpretable reasons for their predictions. In industrial AI, a recommendation without reasoning is a non-starter.
- Embed Privacy Protections. Audit data pipelines to ensure worker or customer data is not inadvertently exposed or retained.
- Align Governance with NIST AI RMF. Map each of the NIST trustworthiness characteristics to a responsible executive function. Assign ownership.
- Train Operators on AI Ethics. Teach not just how the AI works, but what ethical safeguards are in place. Adoption depends on frontline confidence.

HILLSTRONG GROUP SECURITY

Building Resilient, Explainable AI Systems: The Hard Truth About Trustworthy Autonomy

Designing AI That Stands Up to Scrutiny and Stress

Factories do not forgive fragile systems. A pump that cracks under thermal stress, a sensor that drifts out of calibration, a safety system that fails under real load: all of them eventually cause production losses or worse. The same is true for artificial intelligence. If an AI system cannot withstand scrutiny, stress, and operational reality, it does not belong in critical operations.

The industry has spent decades designing for resilience in mechanical and control systems. Redundancy, fail-safes, preventive maintenance, and human oversight are baked into the DNA of industrial operations. AI must live up to the same standard. Anything less creates a brittle dependency that erodes trust and magnifies risk.

This final piece in the series argues a simple point: resilient AI in industry is not about autonomy at any cost. It is about explainability, testability, and recoverability, the traits that make technology dependable when human lives and billions in assets are at stake.

Why This Matters Right Now

AI adoption in manufacturing, energy, and critical infrastructure has moved beyond pilots. Plants are already relying on AI for predictive maintenance, defect detection, scheduling, and anomaly detection. Boards are allocating budgets on the promise of efficiency and insight. But adoption is hitting the same roadblock everywhere: trust.

Operators hesitate to act on black-box recommendations. Engineers struggle to explain why a model flagged one event but missed another. Cyber teams question whether adversarial manipulation could blind their anomaly detectors.

Resilience and explainability are not academic concerns. They directly determine adoption. If operators cannot understand why AI made a recommendation, they will ignore it. If AI collapses under unexpected conditions, trust is lost. If there are no manual fallback procedures, business continuity is jeopardized.

AI in industry does not just need to be accurate. It needs to be trustworthy under pressure.

The Three Pillars of Industrial AI Resilience

What makes AI actually work in a plant? I've found it comes down to three non-negotiable factors:

- Explainability: The "show your work" principle. When an AI flags a pump for maintenance, operators demand to know why. Is it vibration patterns? Temperature anomalies? Without this context, alerts just become background noise. Good explainability transforms complex algorithms into actionable insights that can be clearly communicated across all organizational levels. Feature attribution and simplified decision trees serve as valuable tools for achieving this transparency.
- Testability: AI systems require the same rigorous validation as physical safety equipment. Every model must undergo comprehensive testing against a spectrum of scenarios including edge cases, sensor anomalies, and unexpected process conditions. Thorough validation against disruption scenarios is essential for identifying potential failure modes before deployment in production environments.
- Recoverability: AI systems must be designed with failure in mind. This means implementing manual fallbacks for every AI-driven process and establishing clear protocols for human intervention. Properly documented fallback procedures should receive the same attention to detail as the AI design itself, ensuring operational continuity even when systems behave unexpectedly.

Flawed Thinking That Undermines Resilience

- "If the accuracy is high, the system is resilient." This dangerous misconception ignores real-world complexity. Models that show 99% accuracy can fail catastrophically when faced with sensor drift, equipment aging, or seasonal variations that weren't in the training data. Without stress testing, high accuracy means nothing.
- "Operators don't need explainability, just alerts." Consider how operators typically handle unexplained alarms: they tend to ignore them. When a supervisor asks "why did you shut down that equipment?" responding with "because the AI said so" doesn't cut it. Context isn't optional, it's essential.
- "Resilience is too expensive." This argument falls apart when considering the financial impact of system failures. A hypothetical manufacturer might save $1 million on resilience testing only to lose $10 million in a single weekend when an AI scheduling system crashes. Penny-wise, pound-foolish.
- "Vendors handle resilience." Organizations should be wary of this assumption. Many contracts promise "99.9% uptime" with fine print excluding common scenarios like power fluctuations or data anomalies. The responsibility ultimately remains with the implementing organization, regardless of what vendor sales pitches claim.

Alternative Perspectives That Shift the Conversation

1. **Treat AI Like Rotating Equipment.** Models should be managed like turbines or pumps: inspected, recalibrated, and retired when performance degrades. AI is not a one-time install. It is a lifecycle asset.
2. **Resilience as Culture, Not Just Code.** Operators need to be trained not only in how to use AI but also in how to challenge it. Building resilience means embedding skepticism and oversight into the culture.
3. **Explainability as a Strategic Advantage.** Organizations that can explain their AI to regulators, auditors, and customers will earn trust faster than competitors. Transparency is not a burden. It is a market differentiator.

4. **Fallback as Design Principle.** Every AI deployment should include manual override and recovery paths. Just as safety systems default to "fail safe," AI systems should default to human authority.

Real-World Flashpoints

- Predictive Maintenance Collapse: A global manufacturer deployed AI to predict gearbox failures. Initial accuracy was impressive, but as seasonal conditions shifted, drift set in. False positives soared. Without retraining schedules or fallback, operators stopped trusting the system. Production reverted to reactive maintenance, and the investment stalled.
- Quality Inspection Black Box: A food processor adopted AI-driven vision systems. The model flagged defective products but could not explain why. Line operators, unsure if the AI was right, bypassed the system. Regulators later questioned the process. The root problem was explainability, not detection.
- Cyber AI Blindness: An energy company relied on AI anomaly detection for OT cybersecurity. Attackers injected subtle adversarial noise into traffic, blinding the system. Because there was no fallback detection path, malicious activity went unnoticed for weeks. The flaw was not in the AI itself but in the lack of recoverability planning.

Each example illustrates the same truth: resilience is not an optional feature. It is the difference between AI that earns trust and AI that gets sidelined.

Practical Actions for Executives

- Mandate Explainability. Require that every AI recommendation comes with interpretable reasoning. If your operators cannot explain the "why," adoption will fail.
- Stress-Test AI Like Equipment. Design validation protocols that mimic abnormal conditions: sensor failures, adversarial noise, process upsets. Test AI resilience, not just accuracy.
- Budget for Lifecycle, Not Launch. Fund retraining, validation, and recalibration as part of ongoing OPEX, not just initial CAPEX. Treat models like living assets.
- Design Fallback Procedures. Ensure manual override is always possible. Document the escalation path when AI recommendations conflict with operator judgment.
- Embed AI Resilience in Culture. Train operators and engineers to treat AI as a tool to be supervised, not a master to be obeyed. Culture, not just code, sustains resilience.

Closing Thought

Resilient, explainable AI is not about perfection. It is about building systems that can be trusted to operate under stress, challenged when they make mistakes, and recovered when they fail. True autonomy in industry will never mean eliminating humans. It will mean equipping humans with AI that enhances judgment, stands up to scrutiny, and fails safely.

About the Author



Roger Hill

CEO, Hillstrong Group Security, Inc.

Roger brings over 30 years of experience in industrial automation and cybersecurity. His extensive background in OT security has positioned him as a trusted advisor to CISOs across various industries. His expertise lies in developing and implementing robust OT GRC programs that address the unique challenges faced by global manufacturing operations. Throughout his career, he has been dedicated to helping organizations achieve operational resilience through strategic security initiatives and pragmatic solutions. His commitment to excellence and deep understanding of the intricacies of OT systems have earned him a reputation as a pioneer in the field. Roger's methodologies continue to drive the development of secure and compliant OT environments for leading manufacturing companies worldwide.

HILLSTRONG GROUP SECURITY

About Hillstrong Group Security, Inc.

Hillstrong Group Security, Inc. is an operational-technology cybersecurity consultancy helping industrial and critical-infrastructure operators build security programs that hold up across their regulatory, product, service, and grid obligations. Blending OT, IoT, and ICS expertise, Hillstrong partners with companies in energy, manufacturing, water, chemicals, food and beverage, automotive, and other essential sectors to achieve operational resilience. Every engagement starts on the plant floor, hands-on and on-site, not on a remote scan.

Our Core Services

- OT Security Maturity Assessments
- OT Risk & Vulnerability Analysis
- OT Program Design & Governance
- Expert-Level OT Security Engineering
- Fractional CISO & Advisory Support
- OT Governance, Risk & Compliance (GRC) as a Service
- OT Staff Augmentation & Contract Specialists
- Resilion.io, AI Execution Engine for OT Cybersecurity

Resilion.io is Hillstrong's AI-powered OT cybersecurity platform, delivering continuous visibility, risk quantification, and program execution (build, execute, manage), so assessment findings translate into measurable resilience.

www.hillstrongsecurity.com · Based in Atlanta, GA · Solving cybersecurity challenges globally