THE AI TRUST GAP IN CYBERSECURITY 2026

Why Security Leaders Embrace AI Assistance but Resist Uncontrolled Autonomy

A Survey of 500 AI and Cybersecurity Decision-Makers

Infopercept

INYINSENSE

Executive **Summary**

Artificial intelligence has moved from experimentation to active adoption in cybersecurity. Security leaders are deploying AI to process security data, enrich threat intelligence, prioritize alerts, accelerate investigations and improve the productivity of security teams. Infopercept' s survey of 500 AI and cybersecurity decision-makers found that

72%

of organizations are already using, piloting or actively evaluating AI for cybersecurity.

But widespread interest in AI does not translate into willingness to give AI unrestricted authority.

The AI Trust Gap

The survey reveals a significant gap between AI adoption and AI trust.

While **81%**

of respondents are comfortable with AI assisting security analysts

only **14%**

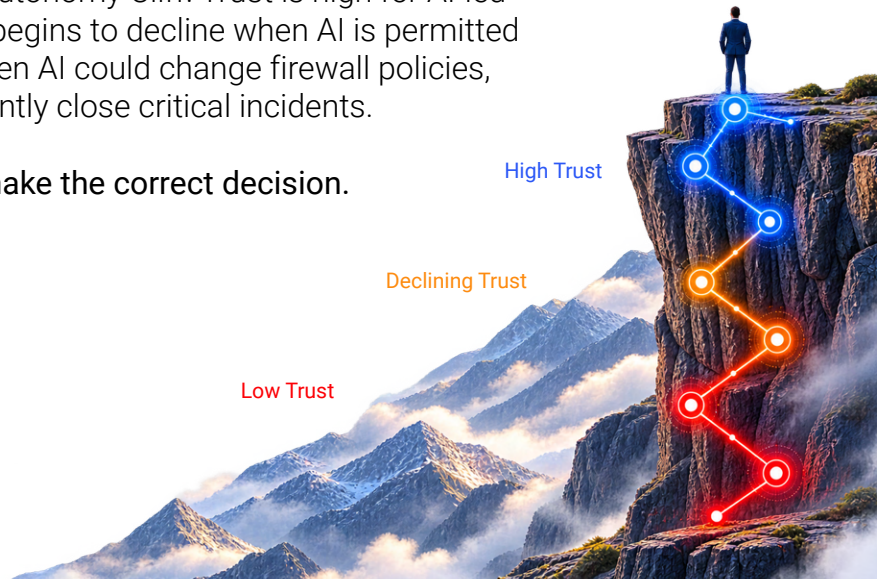
support allowing AI to take unrestricted autonomous action.

Security leaders are broadly willing to use AI to analyze, investigate and recommend. Their confidence declines sharply when AI begins making decisions or executing actions that could affect users, systems or business operations. **This is the AI Trust Gap.**

The Autonomy Cliff

The survey also identifies what we call the Autonomy Cliff. Trust is high for AI-led enrichment, correlation and prioritization. It begins to decline when AI is permitted to block, isolate or disable. It falls further when AI could change firewall policies, remediate production systems or independently close critical incidents.

The concern is not simply whether AI can make the correct decision.



Security leaders are equally concerned about:

- The consequences when AI is wrong
- The ability to explain AI decisions
- Consistency of AI outputs
- Lack of business and organizational context
- Accountability for incorrect actions
- The ability to audit AI decisions

The findings suggest that security leaders are also becoming more sophisticated in how they evaluate "AI-powered" cybersecurity products.

Only

18%

are very confident that they understand the AI technologies operating inside their security stack.

Meanwhile

87%

believe the market either combines general-purpose and purpose-built AI, frequently layers general-purpose AI onto existing tools, or uses the term "AI-powered" too broadly to make the underlying technology clear.

This does not mean security leaders reject general-purpose AI or large language models. These technologies have clear value in information-intensive and language-driven security workflows.

However, the survey suggests that security leaders distinguish between

AI intelligence and AI authority.

For high-consequence decisions,

84%

say they need confidence that similar inputs will produce consistent and controlled outcomes.

48%

Selected
Controlled Autonomy

31%

Preferred
Human-in-the-Loop AI

8%

Preferred
Unrestricted AI Autonomy.

When asked which model of AI autonomy they trusted most, the largest group - 48% - selected Controlled Autonomy, where AI can act independently within predefined policies, permissions and risk boundaries. Another 31% preferred Human-in-the-Loop AI. Only 8% preferred unrestricted AI autonomy.

The message is clear.
Security leaders do not reject autonomous cybersecurity. They reject uncontrolled autonomy.

The future of cybersecurity AI may therefore not belong to systems that simply maximize autonomy.

It may belong to systems that combine:

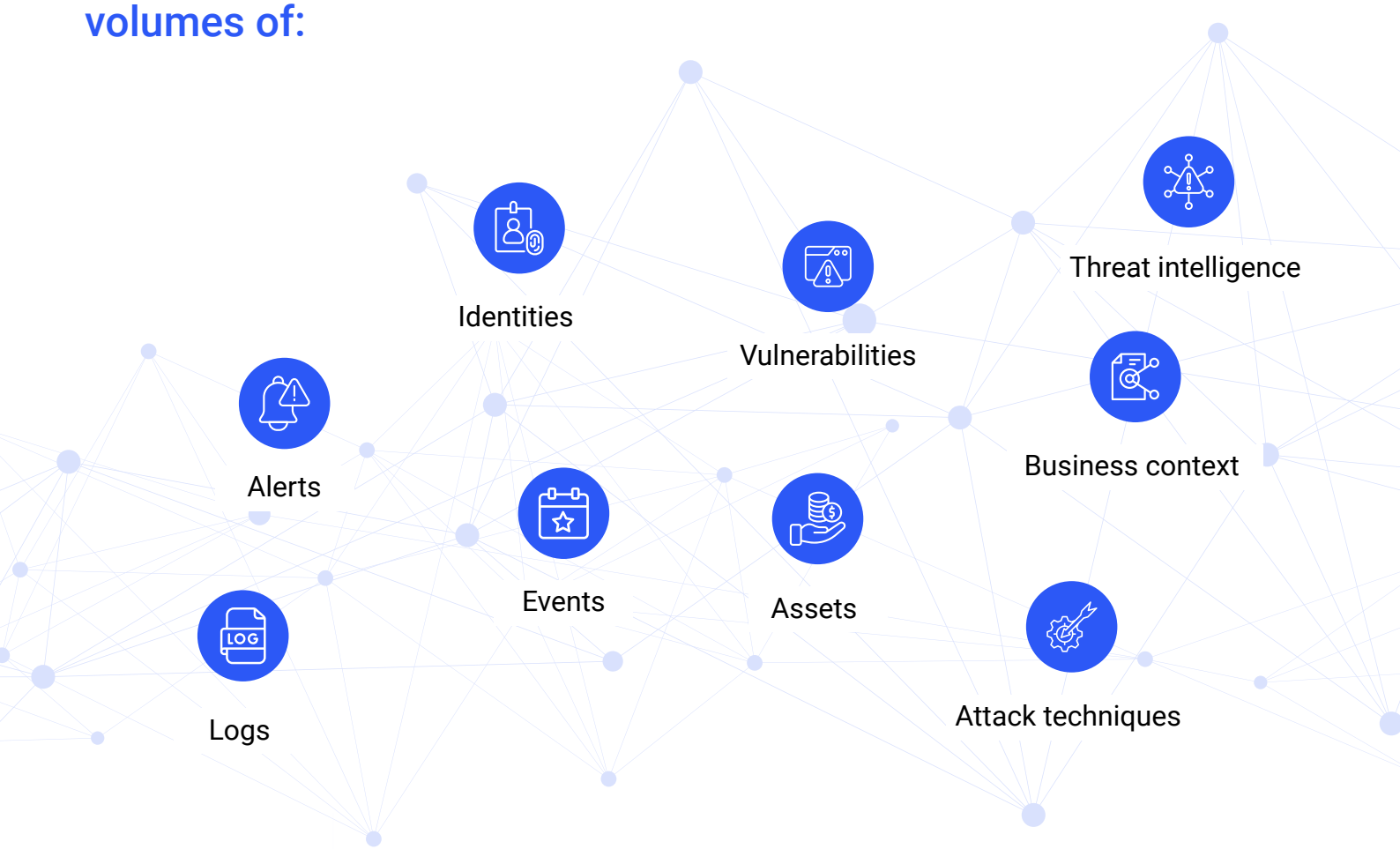
AI Intelligence + Security Context + Deterministic Control

The survey's final conclusion is:
The future of cybersecurity AI is not autonomous by default. It is autonomous by design—within boundaries.

1. Introduction: **AI Has Arrived.** Trust Is Still Catching Up.

Cybersecurity is one of the most natural environments for AI.

Security teams must continuously process enormous volumes of:



The volume and speed of modern cyber threats have created a clear case for machine assistance.

AI

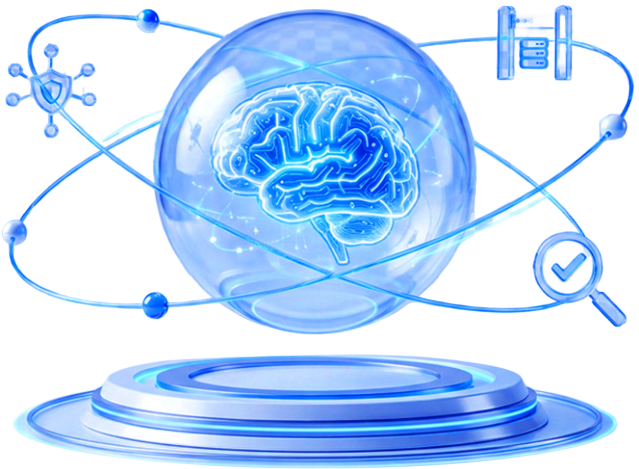
AI can process information faster than humans. It can identify patterns across enormous data sets. It can help investigators understand complex incidents. It can summarize evidence and recommend actions.

The industry is therefore rapidly embracing AI.

Published research has also documented this trend.

ISC2

ISC2's 2025 AI Pulse Survey found significant AI security adoption, testing and evaluation among cybersecurity professionals.



But this survey began with a different question.

Not:

Is AI useful for cybersecurity?

Instead:

How much authority are security leaders prepared to give AI?

2. About the Research

Survey Sample

The research surveyed **500 decision-makers involved in AI and cybersecurity decisions.**

Respondents included:

- CISOs
- Deputy CISOs
- CIOs and CTOs with security responsibility
- Heads of Cybersecurity
- Security Operations leaders
- SOC leaders
- Security and Enterprise Architects
- AI and AI Security leaders
- Technology Risk and GRC leaders

All respondents had direct involvement in evaluating, selecting, approving or governing cybersecurity and/or AI technologies.

Organization Profile

The survey focused on medium-to-large and large enterprises where cybersecurity decisions can have material business and operational consequences.

35%

1,000–4,999 Employees

65%

5,000–9,999 Employees

Industry Representation

The sample included respondents from:

BFSI

Technology and IT Services

Manufacturing

Healthcare and Pharma

Telecom

Energy and Utilities

Government and Public Sector

Retail and E-commerce

Other major industries

3. Key Findings at a Glance

72%

Are already using, piloting or actively evaluating AI for cybersecurity.

81%

Are comfortable with AI assisting security analysts.

76%

Are comfortable using AI to support security investigations.

14%

Support unrestricted AI autonomy.

84%

Need consistent and controlled outcomes for high-consequence cybersecurity decisions.

71%

Say business disruption from an incorrect AI decision is a major concern.

67%

Are concerned about the explainability of AI decisions.

63%

Are concerned about inconsistent or unpredictable AI outputs.

84% Want greater transparency or see a mix of general-purpose and purpose-built AI behind today's "AI-powered" cybersecurity products.

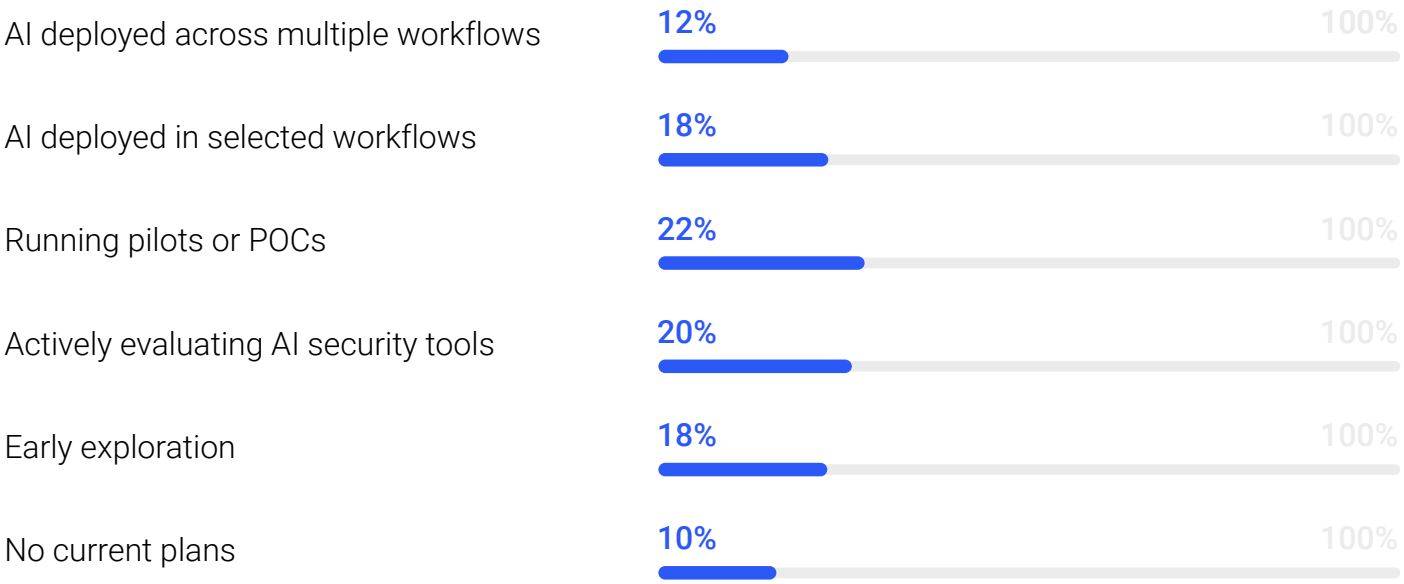
79% Prefer Controlled Autonomy or Human-in-the-Loop AI over unrestricted autonomy.

Finding One: AI Has Entered **Security Operations**

72%

72% are already using, piloting or evaluating AI
The survey confirms that AI is no longer a distant cybersecurity capability.

Survey Sample



Together, **72%** of respondents are already using, piloting or actively evaluating AI. The figure is directionally consistent with external research showing that cybersecurity organizations are moving beyond simple awareness toward deployment, testing and evaluation.

However, adoption should not be confused with maturity.

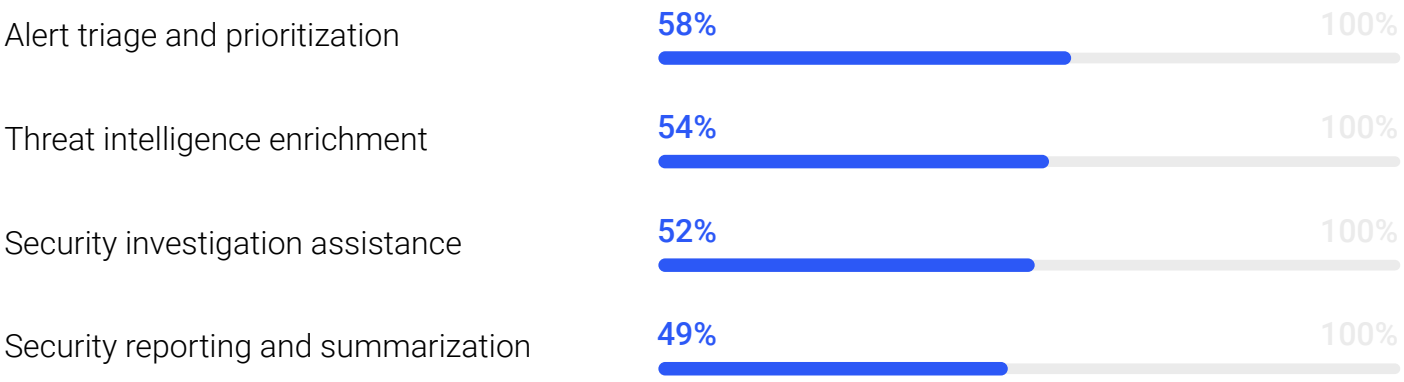
30%

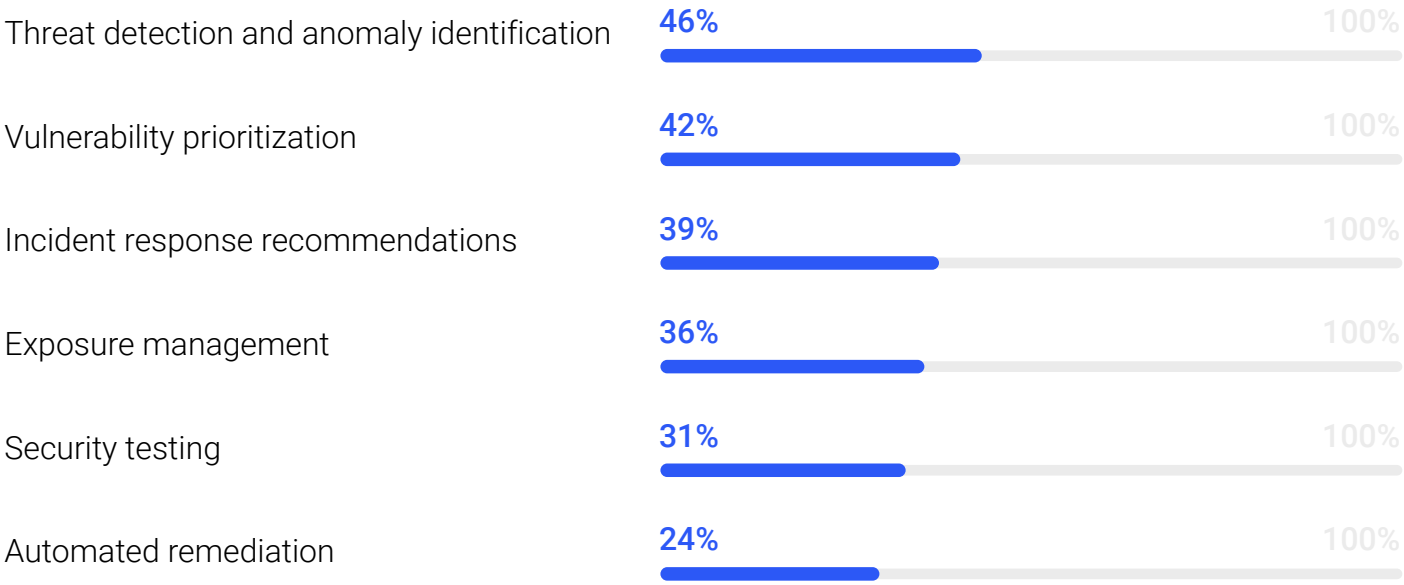
Only 30% currently have AI deployed in one or more security workflows.

For most organizations, the market remains in an important transition:
From experimenting with AI to deciding how much authority AI should receive.

Where AI Is Being Used

The most common AI use cases remain heavily focused on information processing and decision support.



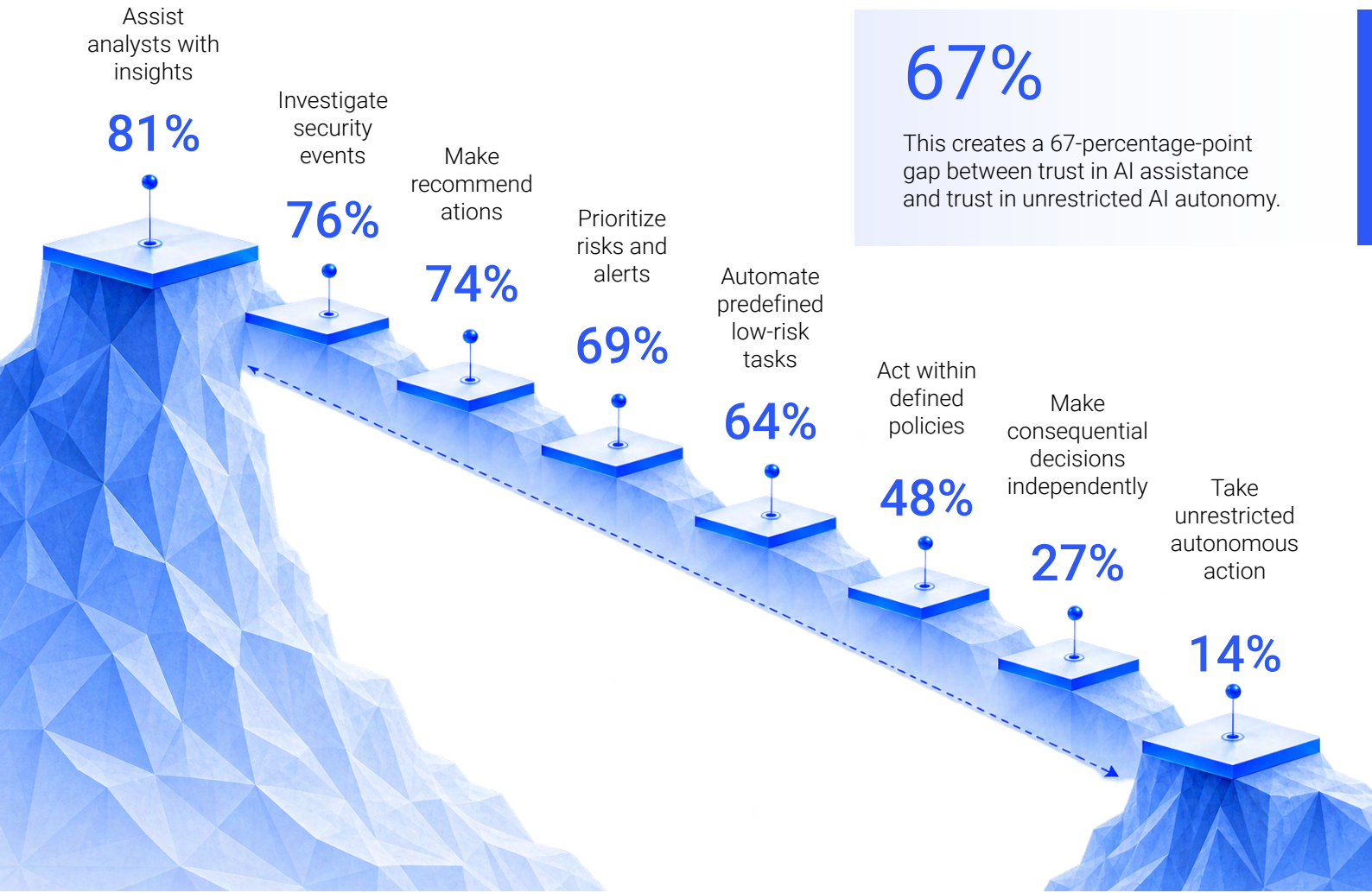


The pattern is significant. The highest-ranking applications help security teams understand and process information. The lowest-ranking application—automated remediation - changes the environment. This distinction appears repeatedly throughout the survey.

4. Finding Two: The **AI Authority Gap**

81% trust AI to assist. Only 14% trust unrestricted autonomy. AI assistance has broad acceptance.

Respondents comfortable with AI performing each role



67%

This creates a 67-percentage-point gap between trust in AI assistance and trust in unrestricted AI autonomy.

We call this:

5. The AI Authority Gap

Security leaders are broadly willing to let AI contribute intelligence. They are much less willing to let AI independently exercise authority. This finding closely reflects the broader industry picture. Darktrace's 2025 survey found that only 14% of cybersecurity professionals allowed AI to take autonomous action without human approval.

The implication is important.

The cybersecurity market should not interpret enthusiasm for AI as automatic approval for autonomous AI agents.

AI adoption is increasing faster than AI authority.

6. Finding Three: The Autonomy Cliff

The survey asked respondents how comfortable they were allowing AI to act autonomously or conditionally across specific security activities.

The results reveal a sharp decline in trust as potential business consequences increase.

Threat intelligence enrichment	78%
Security event correlation	73%
Alert prioritization	68%
Investigation support	62%
Closing low-risk false positives	55%
Blocking known malicious indicators	49%
Isolating a non-critical endpoint	42%
Disabling a standard user account	35%
Disabling a privileged account	24%
Changing firewall policies	22%
Applying remediation to production	18%
Closing a high-severity incident autonomously	16%

The pattern is unmistakable.

Security leaders are most comfortable allowing AI to:

78%	73%	68%	62%
Understand	Correlate	Prioritize	Investigate

Confidence then begins to decline when AI starts affecting the environment.

49%	42%	35%
Block	Isolate	Disable

And it reaches its lowest point for consequential production actions.

22%	18%	16%
Change firewall policies	Remediate production	Close a critical incident

Why does the cliff exist?

The answer is consequence. An incorrect enrichment or prioritization can be corrected relatively easily.

An incorrect production action can create:

Service disruption

Customer impact

Application downtime

Lost productivity

Financial consequences

The survey therefore suggests a fundamental principle for cybersecurity AI:

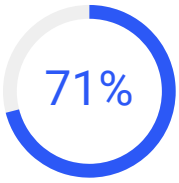
The greater the consequence of being wrong, the greater the demand for control.

7. Finding Four: **The Consequences** of Being Wrong Matter More Than the Speed of Being Right

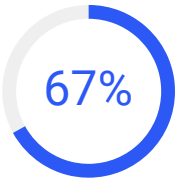
When asked about concerns surrounding AI making consequential security decisions, the leading concern was not AI adoption itself.

It was what happens when AI makes the wrong decision.

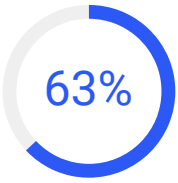
Major concerns



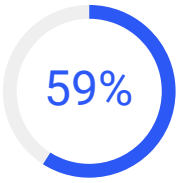
Wrong decision could disrupt business operations



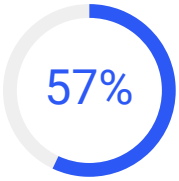
Difficult to explain why AI made a decision



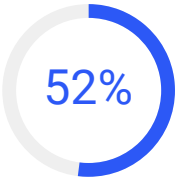
Outputs may be inconsistent or unpredictable



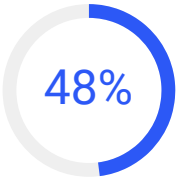
AI lacks sufficient business or organizational context



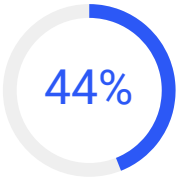
Unclear accountability for a wrong decision



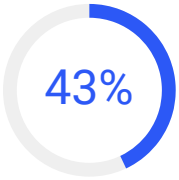
Difficult to audit the decision afterwards



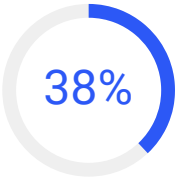
Hallucinations or inaccurate conclusions



Adversarial manipulation



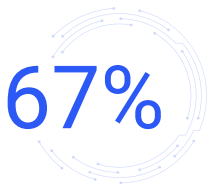
Data privacy concerns



Dependence on third-party AI models

Major concerns

Explainability emerged as one of the strongest barriers to trust.



said difficulty in explaining AI decisions was a major concern.

This is aligned with external findings.

92% Darktrace reported that 92% of cybersecurity professionals wanted to understand how defensive AI made decisions before they could trust it

74% while 74% said they were limiting autonomy until explainability improved.

Security leaders do not simply want an AI system to tell them:

> **"This is malicious."**

They increasingly want to know:

Why?

What evidence led to the conclusion?

What context was considered?

What policy authorized the response?

Why was this action selected?

Can the decision be audited?

Can the action be reversed?

This represents a shift from AI output to AI accountability.

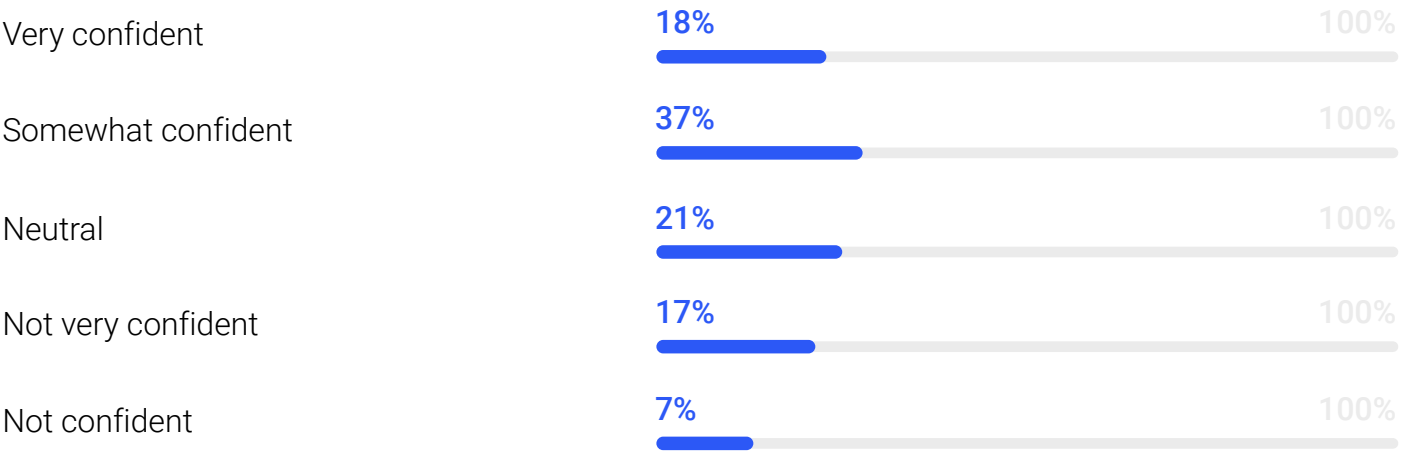
8. Finding Five: The **AI Transparency Problem**

Only 18% are very confident they understand the AI inside their security stack

AI is becoming more visible in cybersecurity products.

But the underlying technology is often less visible.

When asked how confident they were in their understanding of the AI used by their current security tools:

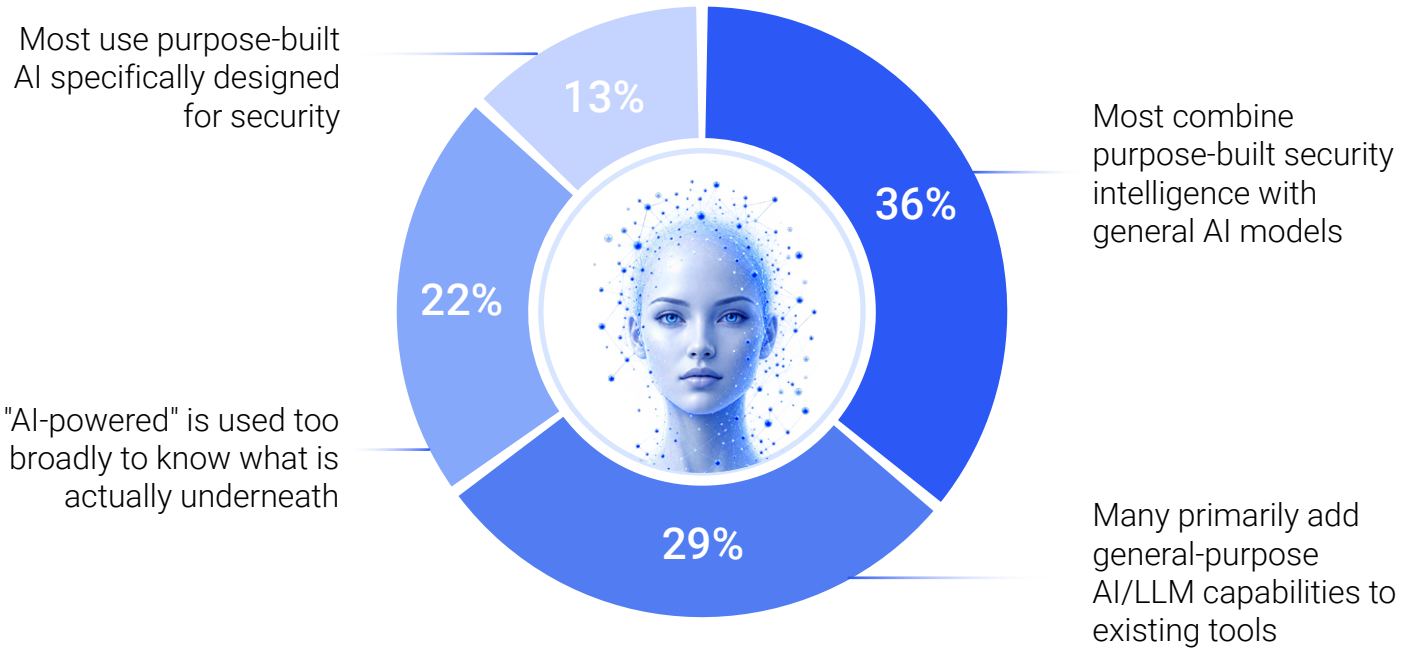


Only 55% expressed confidence. The remaining 45% were neutral or lacked confidence.

Darktrace's research similarly identified a knowledge gap around the types of AI deployed within security environments.

What is actually behind "AI-powered"?

Respondents were asked which statement most closely reflected their perception of today's AI cybersecurity market.



The survey does not suggest that general-purpose AI has no value in cybersecurity.

On the contrary, it can be highly valuable for:

However, the findings suggest that security leaders increasingly want greater transparency about the role AI plays inside a product.

13%

Only 13% believe that most AI cybersecurity products are primarily built on purpose-designed AI specifically for security.

The market is therefore moving beyond:

Does this product have AI?

Toward:

What type of AI does it use?

What does the AI actually do?

What can it access?

What decisions can it make?

What actions can it execute?

9. Finding Six: Security Leaders Want **Probabilistic Intelligence but Deterministic Control**

One of the strongest findings of the survey emerged when respondents were asked to react to the statement:

"For high-consequence cybersecurity decisions, I need confidence that similar inputs will produce consistent and controlled outcomes."

Response

84% agree

46%	38%	10%	4%	2%
Strongly agree	Agree	Neutral	Disagree	Strongly disagree

The implication is not that AI itself must become completely deterministic. That would misunderstand the strengths of modern AI systems. AI can be extremely valuable because it can reason across large and complex information environments. But when AI reasoning leads toward consequential action, organizations want stronger controls governing execution.

This creates an emerging architectural model:

Probabilistic Intelligence

AI can:

- Identify patterns
- Analyze complex data
- Interpret language
- Investigate incidents
- Generate hypotheses
- Recommend actions

Deterministic Control

The security architecture should define:

- Permissions
- Policies
- Risk thresholds
- Action boundaries
- Approval requirements
- Escalation rules
- Audit requirements

The survey suggests that security leaders are not looking for a choice between AI and control. They want both.

Let AI reason. Let policy govern execution.

10. Finding Seven: Security Leaders Prefer **Controlled Autonomy**

When respondents were asked which model of AI autonomy they trusted most, a clear preference emerged.

48%	31%	13%	8%
Controlled Autonomy	Human-in-the -Loop	Predefined automation without AI decision-making	Unrestricted AI autonomy

79% prefer Controlled Autonomy or Human-in-the-Loop AI.

Only **8%** prefer unrestricted AI autonomy. This is one of the most significant findings of the survey.

AI can:	But the organization defines:
Analyze	What AI can access
Investigate	What AI can decide
Correlate	What AI can execute
Recommend	Which actions require approval
Prioritize	Which assets are off-limits
Make selected decisions	Which policies cannot be overridden
Execute authorized actions	When humans must be involved

Extreme 1: Manual Security

- Humans approve everything.
- This may provide control but cannot always operate at machine speed.

Outcome
slow
High control

Extreme 2: Unrestricted Autonomy

- Humans approve everything.
- This may provide control but cannot always operate at machine speed.

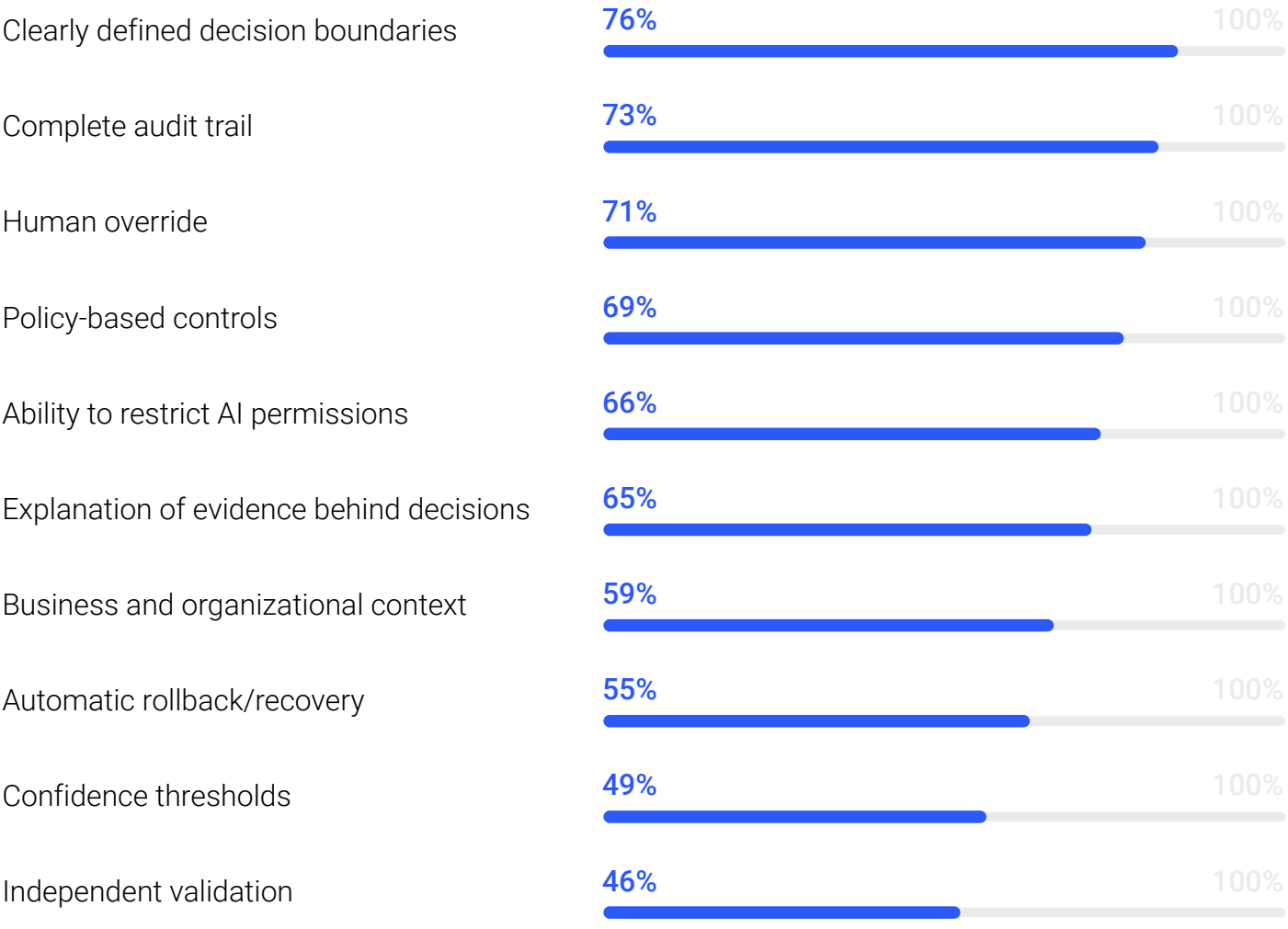
Outcome
Fast
Low control

The Middle Path: Controlled Autonomy

- AI operates at speed.
- But authority is governed.

11. What Would Increase Trust in AI Autonomy?

Respondents were asked which controls would make them more willing to allow AI to take autonomous security action.

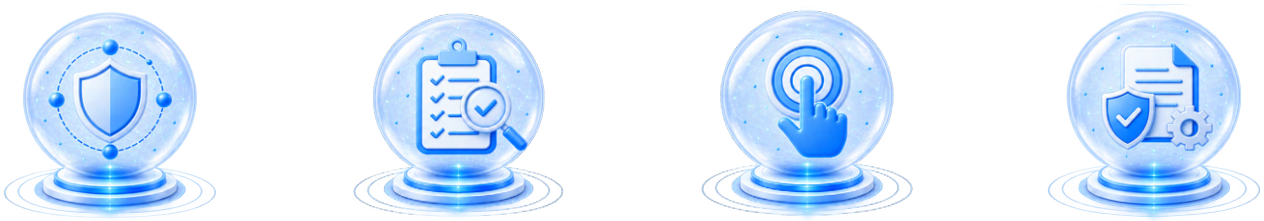


The results are revealing.

Security leaders do not say:

"Make the AI more autonomous."

Their leading requirements are:



Boundaries | Auditability | Override | Policy Control

The survey suggests that AI trust is created as much through governance architecture as through intelligence.

12. The AI Trust Architecture

Respondents were asked which controls would make them more willing to allow AI to take autonomous security action.

1. Intelligence	AI processes information and generates insights.	Question: What can AI understand?
2. Security Context	AI considers assets, identities, threats, vulnerabilities and business criticality.	Question: Does AI understand the environment?
3. Decision	AI recommends or determines an action.	Question: Why did AI reach this conclusion?
4. Policy	Explicit policies evaluate what is permitted.	Question: Is the proposed action allowed?
5. Permission	The AI has only the access required for its authorized role.	Question: Is AI authorized to perform this action?
6. Controlled Execution	The action is executed only within defined boundaries.	Question: Does this action fall within approved authority?
7. Accountability	• Every decision and action can be recorded, reviewed and audited. • Humans retain the ability to define boundaries and override the system	Question: Can we explain what happened?

13. A New AI Maturity Model for Cybersecurity

The survey suggests that AI maturity should not simply be measured by how much AI an organization has deployed.

A more meaningful measure may be:

How much authority has the organization given AI—and how effectively is that authority controlled?

We propose four stages.

Stage 1: AI Assistance

AI provides:

- Insights
- Summaries
- Enrichment
- Analysis

Human authority: Complete.

Stage 2: AI Recommendation

AI proposes actions.

Humans approve consequential actions.

Human authority: Primary.

Stage 3: Controlled Autonomy

AI can act within:

- Policies
- Permissions
- Risk boundaries
- Confidence thresholds

Exceptions escalate.

Human authority: Governance and override.

Stage 4: Unrestricted Autonomy

AI independently decides and executes actions.

Human authority: Oversight.

The Survey Finding

The market is moving toward:

Stage 3: Controlled Autonomy

It is not yet broadly ready for:

Stage 4: Unrestricted Autonomy

14. What Security Leaders Should Do Next

The findings suggest five priorities.

1. Define an AI Authority Model

Clearly establish what AI can:

- Observe
- Analyze
- Recommend
- Decide
- Execute

Do not treat all AI capabilities as equal.

2. Match Autonomy to Consequence

Low-risk activities can receive greater autonomy.

High-consequence actions should require greater controls.

The autonomy granted to AI should reflect the potential impact of an incorrect decision.

3. Ask Vendors What Is Behind the AI

Security buyers should move beyond asking:

"Does your product use AI?"

- They should ask:
- What AI models or technologies are involved?
- What security-specific context does the system use?
- What decisions does AI make?
- Which actions can it execute?
- How are permissions controlled?
- How are decisions explained?
- Can actions be audited?
- Can actions be reversed?

4. Treat AI Governance as Architecture

AI governance should not exist only in policy documents.

It should be embedded into:

- System architecture
- Permissions
- Policies
- Workflow design
- Approval mechanisms
- Logging
- Incident response

5. Measure Trust, Not Just Adoption

AI adoption metrics alone can be misleading.

A security organization may use AI extensively while giving it almost no autonomous authority.

Organizations should therefore track:

- AI usage
- AI accuracy
- AI authority
- AI control
- AI explainability
- AI auditability

15. What **Cybersecurity Vendors** Should Learn

The survey also has important implications for vendors building AI-powered security products.

"AI-powered" is no longer sufficient

The market increasingly wants architectural transparency.

- What intelligence powers the AI
- Where general-purpose AI is used
- Where security-specific intelligence is used
- What AI is allowed to access
- What AI is allowed to decide
- What AI is allowed to execute
- Which deterministic controls govern execution

More Autonomy Is Not Automatically Better

In cybersecurity, unrestricted autonomy can itself become a risk.

The better product question may be:

Unrestricted Autonomy
Higher Risk Lower Control

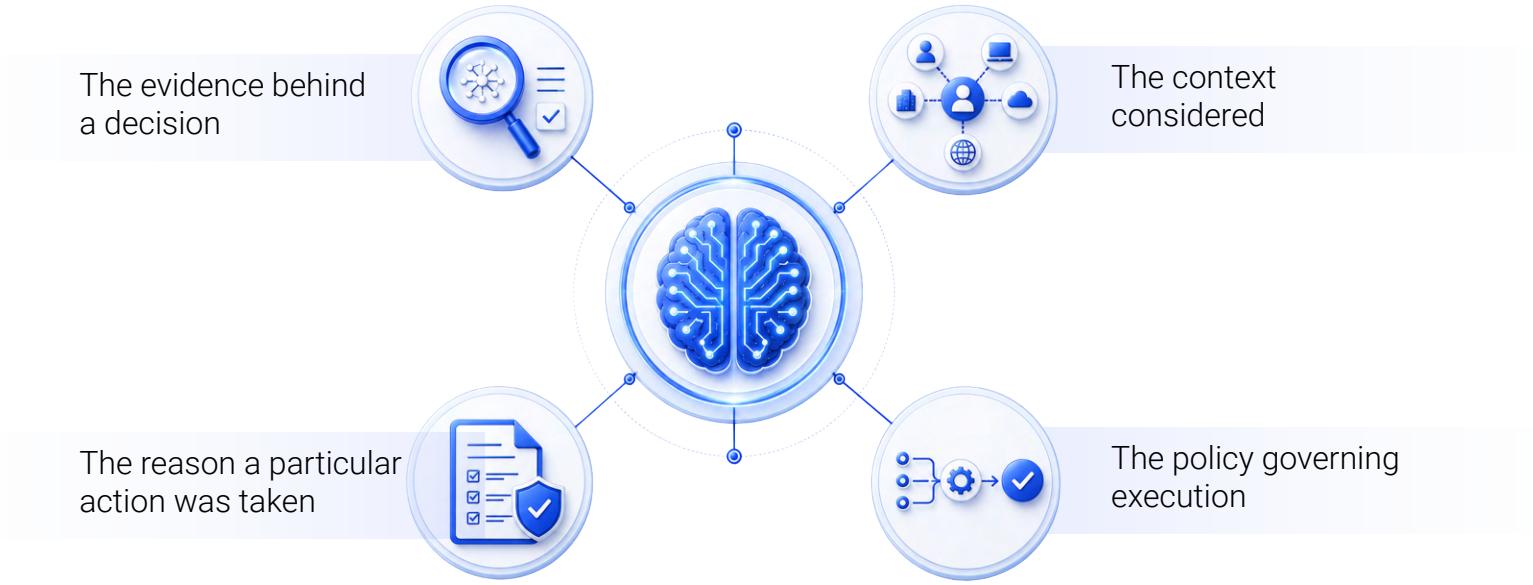
Controlled Autonomy
Balanced Risk Higher Trust

How precisely can AI autonomy be controlled?

Explainability Must Become a Product Feature

AI should not simply produce conclusions.

Security teams should be able to understand:



16. Conclusion: **Intelligence and Authority** Are Not the Same

AI is becoming essential to modern cybersecurity. Security teams need AI to help manage the scale and complexity of the modern attack surface.

But the survey demonstrates that:

- AI intelligence does not automatically earn AI authority.
- Security leaders are comfortable allowing AI to help them understand.
- They are increasingly comfortable allowing AI to investigate.
- They are willing to use AI recommendations.
- But they demand greater controls when AI begins to independently change the security environment.

This is the AI Trust Gap.

The gap is not evidence that AI has failed. It is evidence that cybersecurity has reached the next stage of AI maturity.

The industry is moving beyond:

Can AI help us?

Toward:

Can we control AI well enough to trust it with authority?

The survey's answer is clear. Security leaders do not want to choose between humans and AI. They do not want to choose between intelligence and control.

They want:

Human Authority
to define what matters.

Deterministic Control
to govern what happens next.

AI Intelligence
to operate at machine speed.

The future of cybersecurity AI is therefore unlikely to be:

Autonomy without boundaries.

It will be:

Intelligence within boundaries.

Or, in its simplest form:

Let AI Think. Let Policy Control.

Infopercept

INYINSENSE

Office Address

3rd floor, Optionz Complex
Opp. Hotel Regenta,CG Road,
Navrangpura, Ahmedabad -
380009, Gujarat, INDIA

Contact Detail

www.infopercept.com
sos@infopercept.com
+91 9898857117