CyberOne

WHITE PAPER // CYBERONE SECURITY

The CISO's New Accountability

Governing the Agentic Workforce

An independent industry perspective on why the CISO role structurally changed in the last 18 months — and what Fortune 1000 security leaders must do about it before August 2, 2026.

Authored by

Ricky Allen // Chief Technology Officer, CyberOne Security

Executive Summary

The Chief Information Security Officer role has structurally changed, and most CISOs are operating as if it has not. The old job was defending a perimeter around deterministic systems operated by human employees whose actions could be attributed, audited, and corrected. The new job is governing a non-deterministic workforce of autonomous agents that the CISO did not hire, does not have direct authority over, and cannot fully observe all while remaining personally accountable when those agents cause a reportable incident.

This paper diagnoses the accountability gap facing Fortune 1000 security leaders in 2026, explains why four interlocking problems responsibility, visibility, identity, and evidence cannot be solved with productsalone, and offers a practical framework for closing the gap before regulatory enforcement and board-level scrutiny make improvisation impossible.

> **AI adoption is running at machine speed. Corporate governance is running at human speed. The CISO is the one standing in the gap between the two and right now, the accountability is arriving before the authority.**
>
> *— Ricky Allen, CTO, CyberOne Security*

Key findings in this paper:

- Approximately 90% of organizations have adopted AI in some form. Roughly half have already experienced an AI-related cyber incident.
- In one recent CyberOne Generative AI assessment, 254 unique GenAI applications were in use across 1,150 active users and 42% of users were actively bypassing company AI policy.
- Non-human identities now represent 52% of identity risk, exceeding human users (37%) for the first time in enterprise history.
- 82% of executives are confident their existing policies cover unauthorized agent actions. Only 14.4% of agents are deployed with full security or IT approval. Confidence is running 5x ahead of actual control.
- EU AI Act Annex III obligations for high-risk AI systems take effect August 2, 2026. Penalties reach €15M or 3% of global turnover, whichever is higher.

Introduction: A Different Job, Not an Evolution

Every CISO in a large enterprise has had a version of the same conversation in the last 18 months. The CEO, the board, or a business unit leader asks whether the company is "doing enough with AI." The implication is always that the company needs to move faster. The consequence is always that the security function is handed responsibility for a category of risk that did not exist when the CISO signed the employment contract.

What makes this shift different from previous technology transitions is that it is not incremental. Cloud adoption expanded the attack surface. Mobile expanded the endpoint inventory. Remote work dissolved the perimeter. Agentic AI does something categorically different: it introduces a class of autonomous actors operating inside the enterprise actors that make decisions, take actions, and modify state without direct human supervision. These are not tools. They are delegated workers.

The industry consensus, visible across RSAC 2026, Gartner's 2026 Data & Analytics predictions, and regulatory signals from Brussels, Washington, and London, is that this shift is real and accelerating. What the consensus has not yet clearly articulated is what it means for the personal accountability posture of the CISO.

The framework

The accountability gap is four interlocking problems. Each is a consequence of theone before it. Security leaders who attempt to solve any one of them in isolation will find themselves exposedon the other three.

01
THE GAP
Responsibility
Who owns AI risk?

02
THE BLIND SPOT
Shadow Agents
What can't you see?

03
THE GAP BENEATH
Identity
On whose behalf?

04
THE TEST
Evidence
Can you prove it?

The remainder of this paper examines each theme, quantifies its current state, and recommends concrete actions for Fortune 1000 CISOs preparing for the next 12 to 18 months.

SECTION 01 // THE RESPONSIBILITY GAP

Who owns AI risk?

The organizational problem underneath everything else. AI adoption moves at exponential speed. Corporate governance moves at linear speed. The CISO stands in the gap and carries the accountability for it.

Two Speeds, One Accountability

The speed mismatch between AI adoption and corporate governance is the most cited diagnostic of the current moment, and it is cited correctly. Every function inside the enterprise is pushing to deploy AI. The tools are inexpensive, the business cases are compelling, and the barriers to entry are near zero. A non-technical employee can build a functional AI agent inside Microsoft Copilot Studio before lunch. Corporate governance, by contrast, moves at the pace of policies, committees, board education cycles, and regulatory frameworks.

~90% / ~50%

of organizations have adopted AI in some form. Of those, roughly half have already experienced an AI-related cyber incident.

Fractured Ownership

The reason the gap is so difficult to close is that AI risk does not have a single owner inside most large enterprises. It has five or six partial owners, none of whom carries full accountability. Data science teams own the models. ML operations teams own the deployment pipeline. Product teams own the integration of AI capabilities into customer-facing offerings. Legal owns compliance. Privacy owns data handling. All while security inherits the consequences when any of the above fails and is typically brought in at the end of the process, if at all.

This fragmented ownership is not accidental. It is how most Fortune 1000 organizations structure AI initiatives in order to move quickly. The speed comes from fragmentation. The liability, however, consolidates and lands on the CISO when the regulator arrives, when the 8-K must be filed, and when the board asks whose decision it was.

> **When a regulator shows up asking for an audit trail of autonomous decisions, they do not call the product manager who deployed the feature. They call the person who signed the controls disclosure. That is the CISO.**
>
> *— Ricky Allen, CTO, CyberOne Security*

The Accountability Asymmetry

Three regulatory and legal facts define the current accountability posture for Fortune 1000 CISOs. Each of them is already operative today. Each applies to agent-caused incidents without meaningful carve-outs.

SEC Form 8-K Item 1.05

The SEC's four-day material incident disclosure rule applies to incidents caused by AI agents with no exception. Public companies must file on Form 8-K Item 1.05 within four business days of determining that a cybersecurity incident is material. The materiality determination process is itself a disclosable element under Regulation S-K Item 106.

Personal Liability Precedent

Joe Sullivan, the former Chief Security Officer at Uber, was criminally convicted for his role in concealing a breach. Timothy Brown, then CISO at SolarWinds, was named personally in an SEC civil action. Subsequent judicial rulings have narrowed some aspects of the SEC's reach, but the foundational precedent — that the CISO chair carries personal exposure for cybersecurity governance failures — is established.

EU AI Act (Enforcement: August 2, 2026)

Annex III high-risk AI system obligations take effect approximately 100 days from publication of this paper. Penalties reach €15 million or 3% of worldwide annual turnover, whichever is higher. For any Fortune 1000 enterprise with EU-facing operations or EU data subjects, the probability thatzero AI systems fall under Annex III is effectively zero. Credit decisioning, employment screening, healthcare triage, critical infrastructure operations, and regulatory reporting are all explicitly in scope.

The Uncomfortable Question

These facts produce one question every security leader should be prepared to answer honestly: **Do you have the authority commensurate with the accountability you are being handed?**

For most CISOs, the answer today is no. Accountability has arrived. Authority has not. Closing that gap is not a technology project. It is a board-level conversation about organizational design: the CISO's authority to set AI governance policy, block deployments that cannot meet evidence requirements, and require accountability from every function that deploys agents. Without that authority, the accountability cannot be discharged.

PRACTITIONER'S NOTE

The board conversation to have in 2026 is not "how are we using AI." It is "how are we governing AI, and does our CISO have the authority to enforce the governance decisions the board is about to be held accountable for?" If the answer is unclear today, it will be clarified during a post-incident deposition. That is the wrong venue.

SECTION 02 // SHADOW AGENTS

What can't you see?

The visibility problem. Shadow GenAI was just the beginning. Shadow agents are the actual breach and they are hiding inside the tools your security team has already approved.

Shadow GenAI Was the Warning. Shadow Agents Are the Breach.

Shadow AI or the use of unsanctioned AI tools outside security team visibility is now a well-documented phenomenon. IBM's 2025 Cost of a Data Breach Report found that organizations with high levels of shadow AI incur breach costs averaging $670,000 higher than organizations with low or no shadow AI. One in five organizations has already experienced a breach tied directly to shadow AI usage. Only 37% of organizations have detection or governance policies in place to address it.

What CyberOne Is Seeing in the Field

The industry statistics above describe the shape of the problem. In a recent CyberOne Generative AI security assessment conducted for an enterprise client, we measured the following:

1,150
active users

254
GenAI apps in use

42%
bypassing policy

This assessment was an enterprise with an established AI usage policy. Two hundredand fifty-four distinct generative AI applications were in active use. The security team had visibility into fewer than twenty of them. Legal had not reviewed the data flows of any application outside that short list. Forty-two percent of users were actively bypassing stated policy. We see this pattern consistently across engagements in multiple industriues.

What this data describes is the shadow *GenAI* problem, where the risk vector is data exfiltration. Shadow *agents* are a different category entirely.

GenAI vs. Agents: Why the Distinction Matters

	SHADOW GENAI	SHADOW AGENTS
Pattern	Human pastes data in	**Agent pulls data autonomously**
Persistence	Single interaction	**Persistent, chained actions**
Risk vector	Data exfiltration	**Action execution**

Attribution	Human session	**Deploying user's credentials**
Blast radius	Bounded	**Orders of magnitude larger**

A shadow GenAI tool receives data. A shadow agent takes actions where it deletes records, sends emails,pushes code, modifies database entries, and calls external APIs. Because the agent typically authenticatesas the deploying user, every action it takes is attributed in the audit log to that human. When a shadow agent does something catastrophic, the audit log will show that a human employee did it. Explaining that distinction to a regulator after the fact is not a defensible posture.

Where Agents Are Actually Hiding

Most AI Security Posture Management (AI-SPM) tools in the current market are designed to discover SaaS-layer generative AI use: ChatGPT, Claude, Gemini, Perplexity, Copilot. They scan for registered accounts, OAuth grants, and network traffic to AI domains. That capability is real and necessary. It is also the easy problem. The hard problem is agents built inside tools the security team has already approved.

The five most common hiding places for shadow agents in Fortune 1000 environments, based on CyberOne engagement data:

Microsoft Copilot Studio	Non-technical users build agents with Microsoft 365 connector access. No IT approval required. Inherits the builder's permissions across SharePoint, Outlook, and Teams.
GPT Actions (ChatGPT Enterprise, Claude)	Users configure agents that call external APIs through custom connectors into CRM, data warehouses, and internal systems. Governance configured in a consumer-grade UI.
Cursor and Windsurf agent modes	AI-powered IDEs with agent capabilities that give the AI read-write accessto entire codebases including secrets, environment variables, anddeployment configurations.
Zapier AI / Make.com	Agents built inside approved automation platforms, inheriting whatever OAuth scopes the deploying user has granted across connected SaaS applications.
MCP servers	Tens of thousands of Model Context Protocol servers are now active across enterprise environments, most installed on individual developer machines, most never reviewed.

These agents do not appear as separate SaaS signups. They appear as authorized usage of platforms that have already been approved. They are hiding inside the trust boundary. This is the core of the visibility problem: existing tooling was not designed to find them.

The Confidence Gap

A 2026 survey of more than 900 executives captured the scope of the overconfidence problem. Both of the following statistics describe the same population:

82%

executives confident policies cover agents

14.4%

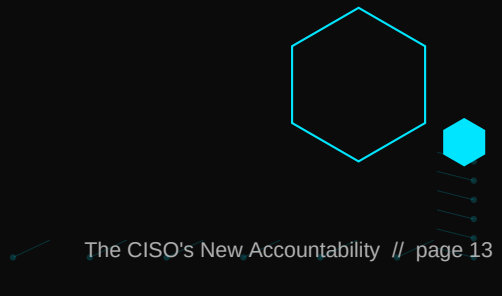
agents with full security approval

Executive confidence is running approximately five times ahead of actual control. This is not a small rounding error. It is a structural mismatch between perception and reality that will manifest in incident reports and regulatory filings over the next 24 months. Security leaders should pressure-test their own confidence against the 14.4% figure before the incident does it for them.

SECTION 03 // IDENTITY

On whose behalf?

The attribution problem. Non-human identities now outrank humans in enterprise risk and no current IAM architecture can reliably answer the question that matters most when an agent takes an action.



The Question Current IAM Cannot Answer

For thirty years, enterprise Identity and Access Management (IAM) and Privileged Access Management (PAM) architectures have been built to answer a single question: *Who triggered this action?* The answer, a user, a service account, an API key producing a principal, a timestamp, an action, and a log entry. This model has anchored cybersecurity audit practice for a generation.

Agents break the model. The question that matters when an agent takes an action is no longer who triggered it: *On whose behalf was the action executed?*

Because when an agent acts, there are typically three or four principals involved: the human who originally deployed the agent, the credentials the agent is using, the service the agent is calling, and often another agent that delegated to this one. Current IAM architectures collapse that chain down to the identity whose token was presented at the API call which is usually the deploying user, who had no awareness the action was taking place.

> **The delegation chain is the foundational gap in enterprise identity architecture for the next five years. Every identity vendor in the market is repositioning toward 'agent identity.' Very few have actually solved the attribution problem.**
>
> *— Ricky Allen, CTO, CyberOne Security*

Non-Human Identities Now Outrank Humans

Tenable's 2026 Cloud and AI Security Risk Report documents a historic inversion in enterprise identity risk. For the first time in enterprise history, non-human identities which include AI agents, service accounts, and machine identities represent the majority of identity risk in Fortune 1000 environments.

52%	65%	49%
NHI identity risk share	NHIs holding ghost secrets	dormant over-privileged identities

Source: Tenable 2026 Cloud and AI Security Risk Report. Non-human identities (52%) now exceed human users (37%) in risk share. 65% of NHIs hold what Tenable terms "ghost secrets" which are credentials that are unused, unrotated, or both. 49% of identities with critical excessive permissions are dormant, meaning they combine two of the most dangerous conditions: elevated privilege and no owner actively monitoring usage.

For most enterprises that have not conducted a dedicated non-human identity audit in the last twelve months, the implication is direct: the majority of your identity risk surface is currently invisible to your security team.

Model Context Protocol: The Integration Backbone That Has Problems

The Model Context Protocol (MCP), originally authored by Anthropic and now adopted across OpenAI, major agent frameworks, and increasing numbers of enterprise SaaS vendors, is rapidly becoming the default integration standard between AI agents and enterprise tools. If your organization has agents in production, it almost certainly has MCP servers in production whether your security team has visibility into them or not.

MCP has serious, documented, and exploited vulnerabilities. We include it here because the governance posture required is impossible without awareness of the actual attack surface.

Vulnerability class	Description
Tool poisoning	Malicious instructions embedded in MCP tool metadata. Academic research published in 2026 found that 5 of 7 major MCP clients have no static validation against this attack.
Rug pulls	MCP tools mutate their own definitions after installation. An organization may approve a tool on day one and have it silently reroute API keys to an attacker-controlled endpoint by day seven.
Session hijacking	Cross-session prompt injection attacks through shared message queues in multi-tenant MCP server deployments.
mcp-server-git CVEs (2026)	Three CVEs in Anthropic's own reference implementation: path validation bypass, unrestricted git operations, and argument injection (CVE-2025-68143, -68144, -68145).
MCPJam inspector RCE	CVSS 9.8 remote code execution. The server listens on all network interfaces (0.0.0.0) with no authentication enabled by default.
DNS rebinding (Vet MCP)	External websites can interact with locally-running MCP servers through the user's browser, bypassing same-origin protections.

CyberOne engagement data suggests that most Fortune 1000 organizations are currently in one of two postures toward MCP: a full moratorium (*"We are not permitting MCP"*), or ungoverned adoption (*"Let engineers use what they need"*). Both postures are defensible for six months yet neither is defendable for two years. The moratorium posture fails because developers route around it and competitors move at machine speed. The ungoverned posture fails because the 2026 headlines, the 8-K filings, the regulatory actions,the post-incident board reviews will all come from there. The work required is a defendable middle path, built around evidence requirements and specific acceptable-use criteria.

SECTION 04 // EVIDENCE

Can you prove it?

The regulator problem. Governance without evidence is theater. Regulators are about to stop accepting theater and the discipline you need already exists in your organization.

Governance Without Evidence Is Theater

The distinction between AI governance policy and AI governance evidence has not yet fully landed in most Fortune 1000 organizations. Policies are widely in place. Governance committees have been established. Board-level discussions of AI risk are now standard. What is typically missing is the evidence chain that demonstrates to regulators, auditors, and litigators that the governance is actually operating. Policy without evidence is no longer acceptable.

What Regulators Will Actually Demand

01 **Complete agent inventory**

Not a sample. Not the agents you know about. Every agent in production, including those built inside tools already approved.

02 **Tamper-evident logs**

EU AI Act Article 12 requires automatic logging at every decision point. Application logs that can be silently altered have zero evidentiary value in a regulatory proceeding.

03 **Reconstructible decision trails**

For any agent action: inputs received, tools called, intermediate decisions made, outputs produced, and humans who reviewed or overrode.

04 **Continuous risk management evidence**

Not a point-in-time assessment. Not a report generated last quarter. Evidence that risk is being actively monitored across the full lifecycle of each agent.

05 **Real human oversight**

Bank of England (February 2026): literal "human in the loop" will inevitably fail at scale. Checkbox HITL will not survive the next cycle of enforcement.

THE TESTABLE QUESTION

If a regulator walked in tomorrow and asked you to reconstruct the decision trail of every agent that touched regulated data in the last 90 days could you? In conversations with CISOs across CyberOne's 2026 engagements, the honest answer is almost always no. "We would need to pull that together" is the wrong answer. Once a regulator is asking, there is no time to pull ittogether. You either have the evidence chain or you do not.

The Exposure Management Lens

The central practical insight of this paper is that Fortune 1000 CISOs do not need to invent a new discipline to govern the agentic workforce. The discipline already exists. It is the same one security teams have applied to vulnerabilities for the last two decades: **exposure management.**

Vulnerability management asks: what CVEs exist in my environment, which ones matter to my business, what am I doing about them, and can I produce evidence of remediation? Agent exposure management asks the same questions with a different asset class: what agents exist in my environment, which ones have access to what, how are they behaving, and can I produce evidence of governance?

VULNERABILITY MANAGEMENT	AGENT EXPOSURE MANAGEMENT
Asset inventory	**Agent inventory**
Known CVEs and exposure database	**Known agent behaviors and decision patterns**
Prioritization by business risk	**Prioritization by blast radius**
Continuous scanning	**Continuous behavioral monitoring**
Evidence of remediation	**Evidence of governance**

This insight changes the operational posture. Security teams do not need to stand up new frameworks, hire specialists in unfamiliar disciplines, or wait for a new category of tooling to mature. They need to apply a mature discipline to a new asset class and to do so before the regulatory deadlines make improvisation impossible.

> **The discipline is not new. The asset class is. Exposure management has been the right answer for cyber risk for twenty years. It is the right answer for agent risk today provided your program actually inventories agents, not just models.**
>
> *— Ricky Allen, CTO, CyberOne Security*

The Vendor Landscape: Products Are Inputs, Not Answers

No product category, and no combination of product categories, closes the accountability gap on its own. The accountability gap is organizational. Products address the visibility gap, the identity gap, and the evidence gap — each of which is necessary but not sufficient. The vendors listed below are representative examples of their categories, not endorsements. Selection decisions for Fortune 1000 security programs should be driven by environmental fit, existing integration footprint, and evidence requirements — not by market position.

Shadow AI / Agent Discovery

Discovers AI tools and agents through SaaS usage telemetry, OAuth grants, and network traffic.

Representative examples: *Opsin Security, Harmonic Security, Reco. Traditional CASB/SSE vendors (Netskope,Cato, Palo Alto Prisma) and endpoint vendors (CrowdStrike, SentinelOne).*

AI Security Posture Management (AI-SPM)

Inventories AI models, datasets, and training pipelines across cloud environments.

Representative examples: *Wiz, SentinelOne, Cyera, Grip, Orca Security, Prisma Cloud, Cloudflare, and Varonis. Most are model-centric rather than agent-centric.*

Runtime Agent Governance / Gateways

Sits between agents and tool calls to evaluate requests against policy and block high-risk actions.

Representative examples: *Prisma AIRS, Zenity (strongest in low-code agent governance), Aim Security, Lakera, Pangea, CalypsoAI.*

Non-Human Identity Governance

Discovers, classifies, and governs service accounts, API keys, tokens, and agent identities.

Representative examples: *Astrix Security, Oasis Security, Clutch Security, Entro, GitGuardian.*

Exposure Management Platforms

Inventory assets (increasingly including AI assets), prioritize by business risk, provide evidence chains.

Representative examples: *Tenable (Tenable AI Exposure, Hexa AI), Qualys, Rapid7. CyberOne is a Tenable partner.*

Evidence and Audit Infrastructure

Provides tamper-evident logging, decision trail reconstruction, and continuous compliance evidence.

Representative examples: *Credo AI, Cranium, Holistic AI, Monitaur. No vendor has yet been tested in a live regulatory proceeding under EU AI Act Annex III.*

GUIDANCE FOR SELECTION

The most common mistake Fortune 1000 security teams make in this category is buying from a single vendor category and declaring the problem solved. The four gaps require coverage across discovery, identity, runtime governance, and evidence. Coverage does not require a single vendor. It requires that the four functions are genuinely operational, tested in tabletop exercises, and producing defensible evidence. Ask the board question first: do we have the authority to enforce the governance outcomes the tools will surface? If the answer is no, the tools will generate findings that cannot be acted on.

Four Questions for Monday Morning

Each theme in this paper maps to a single testable question. Fortune 1000 security leaders who can answer yes to all four are ahead of almost every peer. Those who cannot answer yes have a clear starting point.

01

Do I have a list of every agent operating in my environment, including agents built inside tools my team has already approved such as Copilot Studio,GPT Actions, Cursor, Zapier, MCP servers? Not every AI tool. Every agent.If I cannot produce that list in under a week, that is where the work begins.

INVENTORY

02

Can I answer "on whose behalf" for every agent action in my environment? Or am I still attributing agent behavior to the deploying user's credentials? If the latter, the attribution chain is broken and the audit logs will mislead us when an incident occurs.

IDENTITY

03

Can I reconstruct a decision trail if a regulator asks? Are my logs tamper-evident? Do I have continuous risk management evidence but not point-in-time assessments? If the answer is "we would need to pull it together," the compliance posture is theater.

EVIDENCE

04

Do I have the organizational mandate to govern this? Or am I holding the accountability without the levers to discharge it? If the latter, the board conversation is the single most important action item of the year.

AUTHORITY

Yes to all four: you are ahead. No to any of them: you know where to start.

Getting Started: The CyberOne Generative AI Security Assessment

For Fortune 1000 security leaders who have read to this point and recognize the accountability gap described in their own environments, the most useful starting point is a structured Generative AI Security Assessment. This is the engagement in which CyberOne measured the statistics cited earlier in this paper: 1,150 users, 254 GenAI applications, 42% policy bypass.

What the Assessment Covers

AI Usage Discovery

Identification of generative AI applications and emerging agentic workloads across your environment, including SaaS-layer tools and agents embedded inside approved platforms (Copilot Studio, GPT Actions, MCP servers, low-code automation).

Policy Conformance Measurement

Quantification of the gap between stated AI policy and observed user behavior. This is the measurement that produced the 42% bypass statistic cited earlier.

Non-Human Identity Audit

Discovery of service accounts, OAuth tokens, and machine identities associated with AI deployments including dormant credentials and ghost secrets.

Governance Maturity Assessment

Evaluation of your current AI governance framework against NIST AI RMF, ISO/IEC 42001, and EU AI Act Article 9 requirements. Identification of the specific evidence gaps that would surface in a regulatory proceeding.

Roadmap and Prioritization

A practical 30/60/90-day action plan aligned to your risk profile and regulatory exposure.

Why CyberOne

CyberOne is a vendor-neutral cybersecurity services firm serving Fortune 1000 enterprises across energy, healthcare, financial services, retail, critical infrastructure, and state and local government. Our role is to help security leaders translate the exposure management discipline they already know into the agentic asset class and to produce the evidence chains that regulators will demand.

TO REQUEST A GENERATIVE AI SECURITY ASSESSMENT

Visit cyberonesecurity.com/services/ai-security or contact our team at info@cyberonesecurity.com or (800) 842-8914. Initial consultations are complimentary and scoped to your environment.

Closing: The Accountability Has Already Shifted

The framing of this paper is deliberate. The CISO role did not evolve in the last 18 months. It structurally changed. The accountability posture of Fortune 1000 security leaders is fundamentally different from what it was in 2024, and it will be different again by 2027. What matters now is whether individual security leaders recognize the shift in time to shape how their organizations respond or whether they discover it the way most organizations discover such things, which is through an incident, a regulator, or a deposition.

The CISOs who can answer the four questions in this paper, inventory, identity, evidence, and authority will define what "defensible" looks like for agentic AI governance in the next decade. The CISOs who cannotwill be the case studies that inform the next generation of regulatory frameworks. There is still time to be in the first group.

> **The accountability has already shifted. The only question remaining is whether the CISO community shifts with it.**
>
> *— Ricky Allen, CTO, CyberOne Security*

About CyberOne

CyberOne is a custom cybersecurity solutions firm providing Advisory, AI Security, Cloud Security, Network Security, Offensive Security, Exposure Management, and Staff Augmentation services to enterprise customers. Our Defendable Network™ framework serves as the foundation for every engagement, combining industry-leading products with specialized professional services.

Contact	cyberonesecurity.com // info@cyberonesecurity.com // (800) 842-8914
Headquarters	6851 Communications Parkway, Plano, TX 75024
Publication	Spring 2026