# Managing AI Risk – What Every Business Needs to Know


**The Hartford**

**Jennifer Birnbaum**, Cyber Underwriting Officer

**Daniel Silverman**, Underwriting Director, Specialized CyberTech

**Keith Tagliaferri**, Director, Cyber Claim Practices

**Arete**

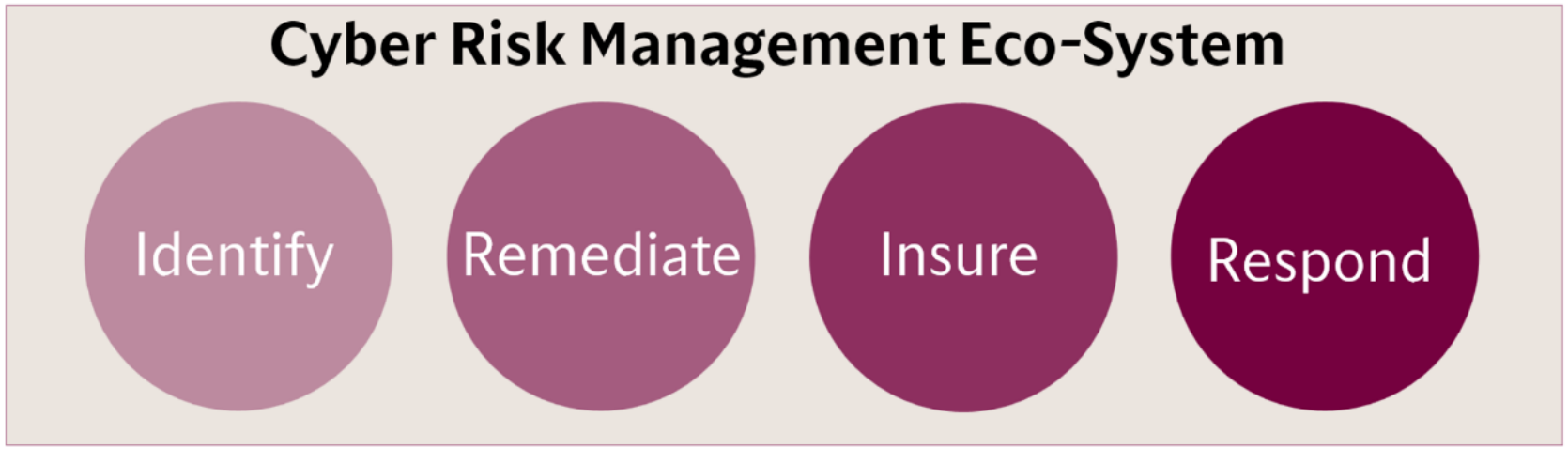
**Zachary Flom**, Associate Director, Threat Intelligence

**Bradley Yandell,** vCiso, Cyber Risk Advisory


**July 30, 2026**

# Our Role – CyberTech Insurance & Risk Management

The Hartford



# Managing Artificial Intelligence Risk

*What Every Business Needs to Know*

Zach Flom, Associate Director, Threat Intelligence
Bradley Yandell, vCiso, Cyber Risk Advisory
28.07.2026

# Agenda

- Emerging AI Risk
- Adversarial Use of AI
- The Foundation for Secure AI Adoption
- The Hartford & Arete Partnership
- Q & A

Founded in
2016

4 Global Offices & Centers
US, Canada, UK & Asia

400+
Employees

Thousands
of Engagements

100s of Partners
Insurance, Law Firms and Technology Partners

1,500+
Managed Security Services Clients

# End-to-End Solutions for Responding to Cyber Threats

## Incident Response

Ransomware Response

Restoration

Business Email Compromise

Threat Actor Communications, Compliance, and Ransom Payment Facilitation

Data Mining and Document Review

## Managed Security Services

Infrastructure and Vulnerability Management

Incident Response Readiness

Cloud Security Solutions

Managed Detection and Response (MDR)

- Arete Bloktd℠ - custom threat detection rules

## Advisory Services

Risk and Compliance

Forensic Investigations

Regulatory Compliance Audits

Expert Witness Services

# Emerging AI Risk

# Emerging AI Risk

- Rapid pace of adoption across all industries and by threat actors
- AI is a force multiplier for adversaries and has enabled them to quickly scale their operations
- Improved identification and exploitation of vulnerabilities has drastically reduced the time between disclosure and weaponization
- Attack timelines and development cycles have been similarly compressed
- Shadow AI
    - AI capabilities enabled by default, growing supply chains/dependencies, unsanctioned internal applications
- Sprawling access and excessive permissions
- Lower barrier to entry for aspiring cybercriminals
- AI hallucination leading to false positives/negatives
- “Cognitive offloading”

# Adversarial Use of AI

# Adversarial Use of AI

The Arete Cyber Threat Intelligence (CTI) Team tracks adversarial use of AI using four separate Priority Intelligence Requirements:

1. How are adversaries incorporating AI into their existing workflows?
2. How are adversaries using AI to develop novel capabilities?
3. How are adversaries targeting enterprise implementations of AI?
4. How are adversaries targeting AI supply chains?

# How are adversaries incorporating AI into their existing workflows?

**Orchestration**

- Threat actors are increasingly leveraging AI models to orchestrate and carry out nearly all phases of a traditional intrusion, from Reconnaissance to Impact
- The integration of AI throughout the attack lifecycle has enabled threat actors to significantly scale operations, with a human operator needing only to act in an executive or supervisory role
- Campaigns that would usually require several personnel with diverse skillsets can now ostensibly be conducted by a single individual
- AI agents can iterate through phases of an attack lifecycle much faster than a human, with average breakout times of less than one hour according to some studies

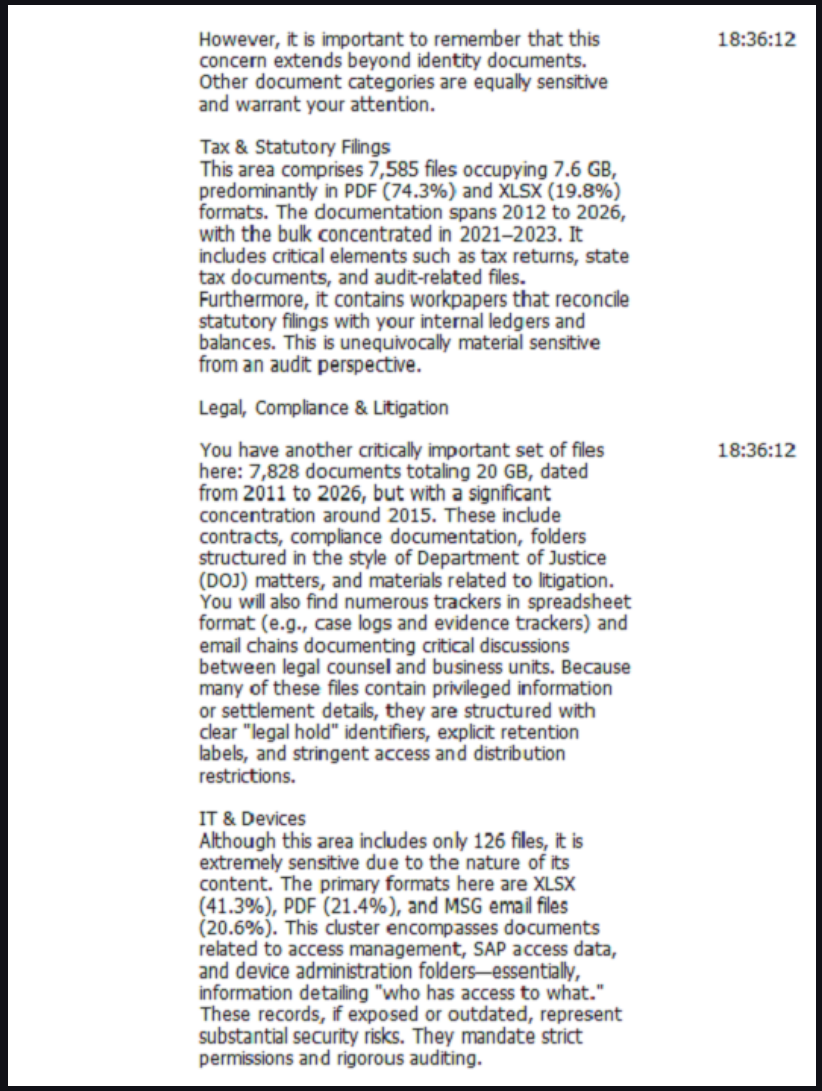
# How are adversaries incorporating AI into their existing workflows?

**Exfiltrated Data Analysis**

- Historically, a bottleneck existed between victim compromise and successful extortion due to limited resources to analyze large volumes of victim data
- LLMs have not only removed this bottleneck, they've made exfiltrated data analysis a core pressure tactic of several ransomware and extortion groups
- Experienced threat actors are now creating standalone Analysis-as-a-Service offerings turning unstructured data sets into finished "intelligence products"
    - Leak Bazaar
- Threat actors observed by Arete leveraging LLMs for data analysis include INC, Settra, and Icarus

**Victim Communication**

- Similarly, threat actors historically have had to dedicate significant time and effort communicating with their victims. Ransomware groups like Ethics and GLOBAL GROUP have incorporated AI chat bots into their negotiation panels. Operators only need to engage directly when negotiations stall or meet certain criteria



# How are adversaries incorporating AI into their existing workflows?

**Capability Development**

- Arete has observed multiple ransomware groups deploying vibe-coded scripts to include Akira and Chaos
- Threat actors rely on LLMs to create new scripts and to modify existing tools at a pace which has not been previously observed
- Threat groups are developing scripts in previously unused languages and/or targeting new technologies
- Hallmarks of LLM-generated code include:
  - Extensive code commenting, more verbose comments
  - Strings or function names in the code indicative of LLM use such as “refined -- reduced false positives”
  - Emphasis on “quality of life” features not typical of malicious scripts

```
1104  # ================================================================================
1105  # 7. FINAL SUMMARY
1106  # ================================================================================
1107  Write-Host "+============================================+" -ForegroundColor Cyan
1108  Write-Host "|            SCAN & DUMP COMPLETE            |" -ForegroundColor Green
1109  Write-Host "+============================================+" -ForegroundColor Cyan
1110  Write-Host "|  Servers scanned  : $($targets.Count)" -ForegroundColor White
1111  Write-Host "|  SQL instances    : $($reachable.Count)" -ForegroundColor White
1112  Write-Host "|  Databases dumped : $(($serverResults.Values | ForEach-Object { $_.Count } | Measure-Object -Sum).Sum)" -ForegroundColor White
1113  Write-Host "|  Export format    : $ExportFormat" -ForegroundColor White
1114  Write-Host "|  Max rows/table   : $MaxRows" -ForegroundColor White
1115  Write-Host "|  Files exported   : $grandTotalFiles" -ForegroundColor White
1116  Write-Host "|  Total rows       : $grandTotalRows" -ForegroundColor White
1117  Write-Host "|  Empty skipped    : $skippedEmpty" -ForegroundColor DarkGray
1118  Write-Host "|  Output           : $exportRoot" -ForegroundColor White
1119  Write-Host "+============================================+" -ForegroundColor Cyan
1120  Write-Host "|  DUMPED DATA BY CATEGORY:                  |" -ForegroundColor Yellow
1121  if ($categoryRows.Count -gt 0) {
1122      $categoryRows.GetEnumerator() | Sort-Object Value -Descending | ForEach-Object {
1123          $catLabel = $_.Key.PadRight(30)
1124          Write-Host ("|    {0}  {1,8} rows" -f $catLabel, $_.Value) -ForegroundColor White
1125      }
1126  } else {
1127      Write-Host "|    (no data exported)" -ForegroundColor DarkGray
1128  }
1129  Write-Host "+============================================+`n" -ForegroundColor Cyan
1130
```

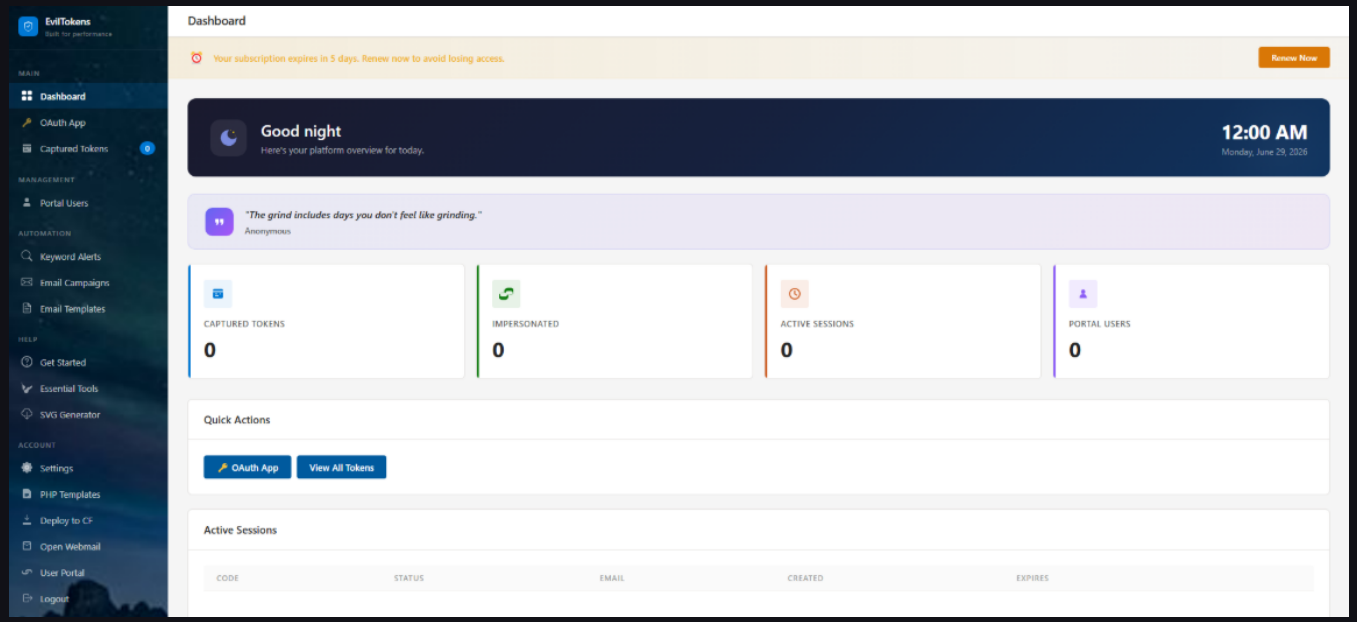
# How are adversaries incorporating AI into their existing workflows?

**Social Engineering**

- AI has enabled adversaries to create far more convincing phishing emails, documents, landing pages, and deepfakes
- Threat actors can now target victims in any language using local and/or industry-specific vernacular, tone, and content
- Information gleaned from open source is used to personalize phishing lures and incorporate current events like conference attendance, M&A, role changes, etc.
- Adversaries use AI to create synthetic personas with detailed backgrounds, social networks, and convincing identification documents

**Phishing-as-a-Service (PhaaS) Kits**

- PhaaS kits provide the above capabilities alongside the infrastructure required to carry out the attack and harvest victim credentials
- Include additional features such as case/resource management, AI chat bot assistants, and pre-built templates



# How are adversaries using AI to develop novel capabilities?

**“Just-in-Time” AI Malware**

- Malware that can modify itself and/or incorporate new functionality by calling LLMs during execution
- Can rewrite its entire codebase, rendering static analysis largely ineffective

**“Vibeware”**

- Threat actors use LLMs to rewrite malware in niche programming languages like Crystal, Zig, and Nim, bypassing most detection rules which are structured based on more prevalent languages

**Purpose-built AI Models and Services**

- Adversaries are creating their own models, “bypass builders”, and data training pipelines tuned for malicious use cases
    - Xantharox AI, WormGPT, KawaiiGPT, DarkBERT, DarkGPT, etc.

# How are adversaries using AI to develop novel capabilities?

**“Slopsquatting”**

- Terminology borrowed from typosquatting tactic
- AI models sometimes hallucinate package names when generating code
- Adversaries create packages with names that match the hallucinated dependencies
- When the LLM hallucinates the package name in the future and the victim searches for the package name, they find the malicious package created by the adversary and load it

# How are adversaries targeting enterprise implementations of AI?

**Prompt Injection and Jailbreaking**

- Threat actors prompt, either directly or indirectly, LLMs to manipulate output, behavior, and to bypass guardrails
- Malicious prompts can be hidden in any content ingested by an LLM like emails, webpages, images, and so on
- Risk is exacerbated by Retrieval Augmented Generation (RAG)

**Data Access and Exfiltration**

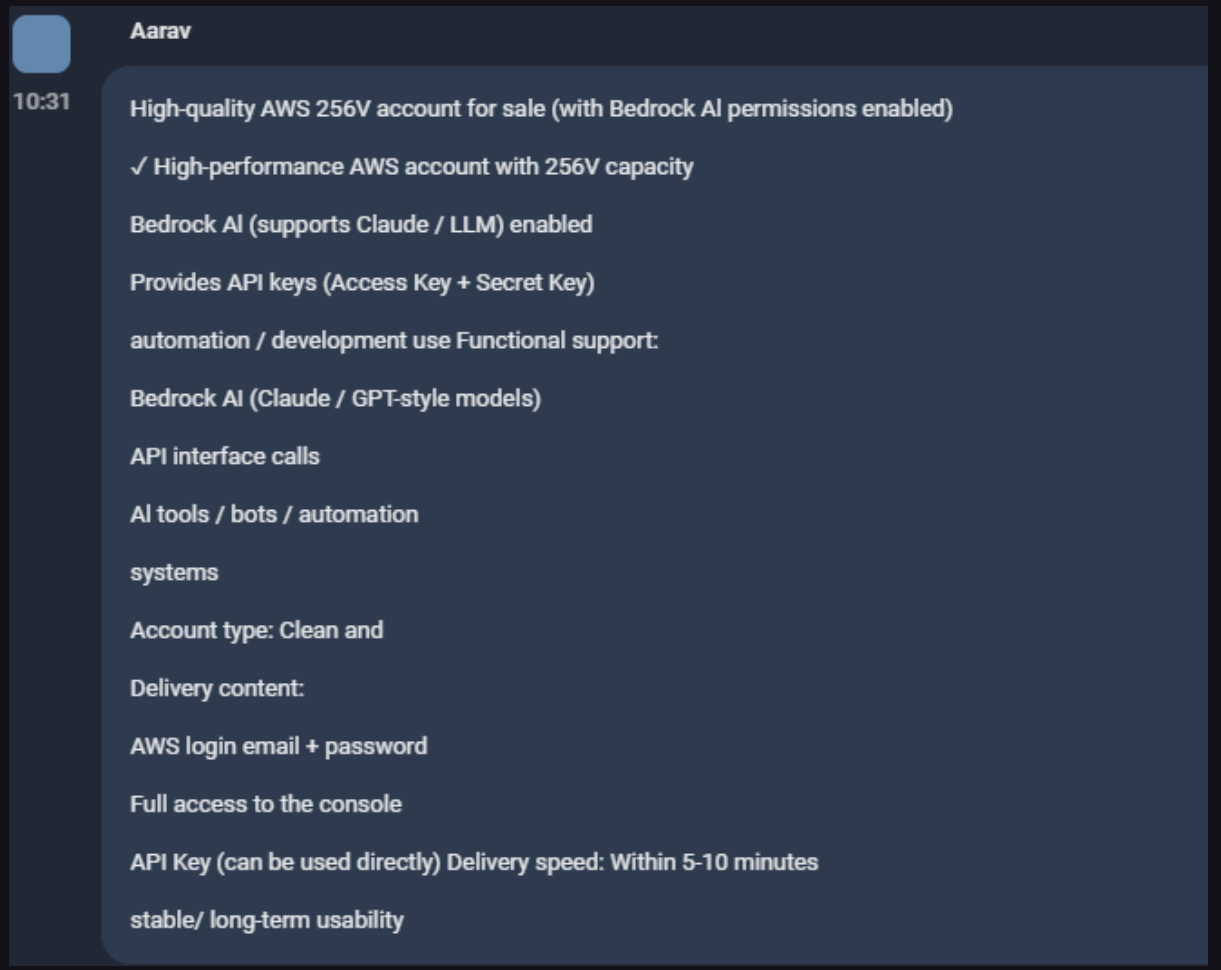
- One of the primary benefits of AI is its ability to synthesize large swaths of data which can be further interrogated for specific insights. In turn, AI agents typically have broad access within an enterprise and elevated permissions
- Threat actors leverage these tools to search for and aggregate sensitive data. They also routinely search AI chatbot logs for sensitive data in previous user sessions


| TimestampDescript | Artifact | Path\URL Accessed |
|---|---|---|
| Target File Last Modified | Jump Lists | C:\Users\[redacted].claude |
| Target File Last Modified | Jump Lists | C:\Users[redacted].claude\.credentials.json |
| Target File Last Modified | Jump Lists | C:\Projects\Repos\[redacted]\CLAUDE.md |
| Target File Last Modified | Jump Lists | C:\Projects\Repos\[redacted]\.claude\settings.local.json |
| Target File Last Modified | Jump Lists | C:\Projects\Repos\[redacted]\.claude |
| Target File Last Modified | Jump Lists | C:\Users\[redacted]\.claude\plugins\config.json |
| Target File Last Modified | Jump Lists | C:\Users\[redacted]\AppData\Roaming\Claude\config.json |
| Target File Last Modified | Jump Lists | C:\Users\[redacted]\.claude.json |
| Accessed | Edge-Internet Explorer | C:\Projects\Repos\[redacted]\CLAUDE.md |

# How are adversaries targeting enterprise implementations of AI?

**LLM Jacking**

- Refers to the unauthorized use of legitimate LLM resources and credentials by nefarious actors
- Adversaries either use the resources themselves or resell access to other actors



# How are adversaries targeting AI supply chains?

**The AI Supply Chain (example)**

- Compute
- Data pipelines
- Code & Open-Source Dependencies
- Foundation Models
- Training/Fine-Tuning
- Model Registries
- Inference APIs
- Agent
- Applications & Users

**Poisoned Software Packages**

- AI tools have complex dependency trees and probabilistic execution which can make it difficult to understand the full extent of the attack surface
- Developer environments are highly targeted due to the presence of valuable credentials with broad, privileged access

# The Foundation for Secure AI Adoption

# Governance & Policies

## AI GOVERNANCE:
### THE FOUNDATION FOR SECURE AI ADOPTION

AI is a powerful enabler—**BUT WITHOUT GOVERNANCE,** it becomes unmanaged business risk.

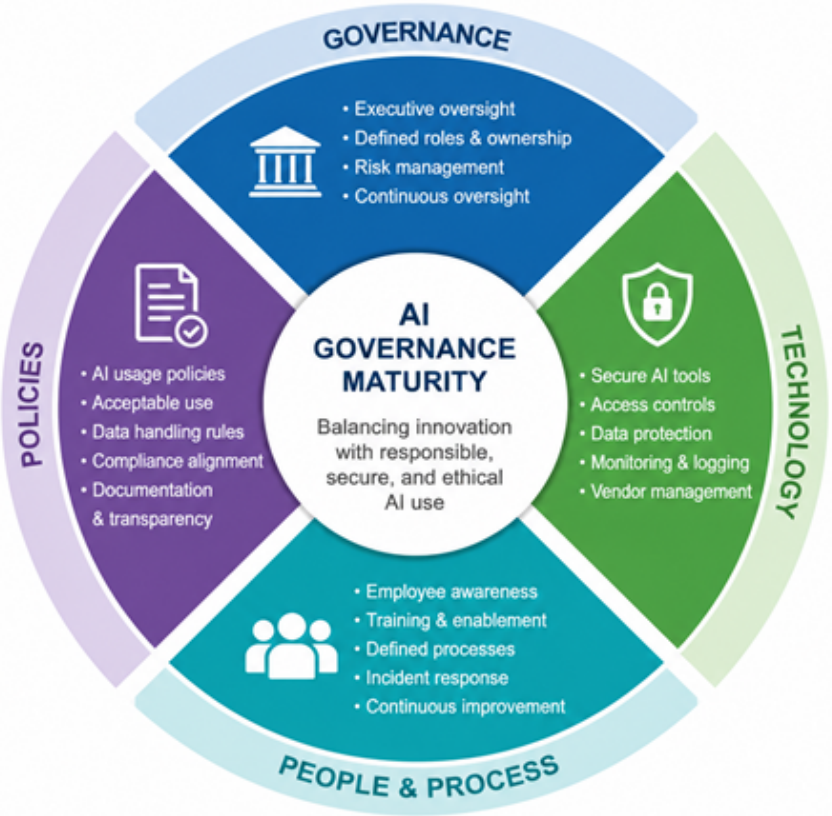
### WITHOUT GOVERNANCE

- Shadow AI and unsanctioned tools
- Data leakage and privacy violations
- Regulatory and legal exposure
- Inconsistent policies and decisions
- Increased cyber and operational risk

### WITH GOVERNANCE

- Executive ownership and oversight
- Clear policies and acceptable use
- Defined roles and accountability
- Standards for data handling and security
- Responsible and innovative AI adoption

## AI GOVERNANCE MATURITY
*A Holistic Approach to Managing AI Risk*



# Secure AI Through Visibility & Control

# People: Your Strongest Defense Against AI Threats

## EMERGING AI THREATS

### AI-GENERATED PHISHING
Highly convincing, personalized attacks at scale

### VOICE CLONING
Impersonates executives or trusted contacts

### DEEPFAKE VIDEO
Realistic fake videos used to manipulate or extort

### BUSINESS EMAIL COMPROMISE
AI enhances the accuracy and timing of impersonation

### EXECUTIVE IMPERSONATION
Targets leadership to gain access, money, or data

## THE AI-POWERED ATTACK CHAIN

**1. ATTACKER CREATES AI CONTENT**
Phishing, deepfake, or voice clone

**2. IMPERSONATES A TRUSTED SOURCE**
Exec, coworker, vendor, or partner

**3. DELIVERS THE ATTACK**
Email, call, chat, or video

**4. TARGETS THE HUMAN**
Exploits trust and urgency

**5. HUMAN VERIFIES & STOPS ATTACK**
Awareness, processes, and verification prevent compromise

**Trust should always be verified.**

## BUILD HUMAN RESILIENCE

### AI AWARENESS TRAINING
Educate employees on emerging AI threats and red flags

### PHISHING SIMULATIONS
Test and reinforce identification of AI-powered phishing

### DEEPFAKE EXERCISES
Help employees recognize manipulated audio and video

### VERIFICATION PROCEDURES
Establish "verify before you trust" processes for all requests

### REPORT & RESPOND
Encourage reporting and quick response to threats

# Structured Approach to AI Risk Management

## AI RISK MANAGEMENT LIFECYCLE

**1. ASSESS**
Understand context, risks, and impacts

**2. GOVERN**
Establish policies, roles, and oversight

**3. PROTECT**
Implement controls to manage risks

**4. VALIDATE**
Test, monitor, and evaluate effectiveness

**5. IMPROVE**
Continuously learn and adapt


## ALIGN WITH LEADING FRAMEWORKS

**NIST AI RMF**
Manage AI risks and build trustworthy, responsible, and secure AI systems through a lifecycle approach.

**NIST CSF & 800-53**
Build a strong foundation with cybersecurity frameworks and federal control standards.

**ISO 27001**
Establish and maintain an Information Security Management System (ISMS).

**ISO 42001**
Implement responsible AI management systems and governance practices.

**SOC 2**
Demonstrate trust and security through independent assurance of your controls.

**CIS CONTROLS**
Implement prioritized, actionable security controls to reduce risk and improve resilience.

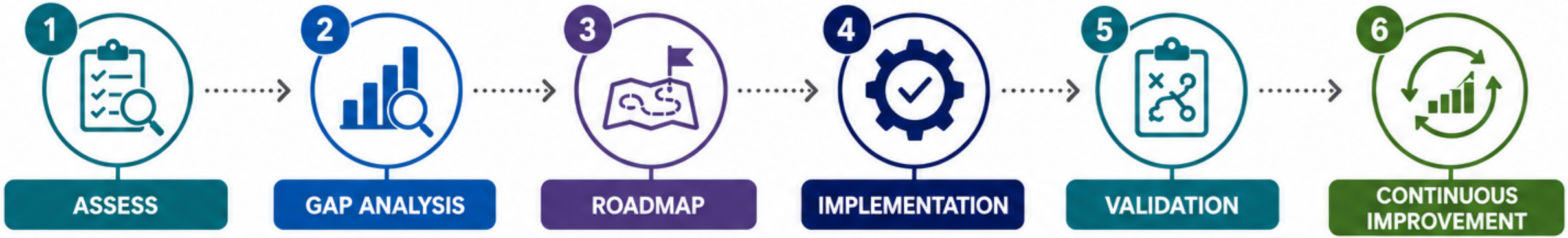
**HIPAA**
Protect patient data and ensure compliance with healthcare privacy and security regulations.

# The Hartford & Arete Partnership

# From Assessment to Resilience

## A trusted partner every step of your AI risk journey.

**We combine deep expertise with a proven methodology to help you govern AI securely and confidently.**



# AI Risk & Compliance Support

## REAL IMPACT. MEASURABLE RESULTS.

**BEFORE**

- Unclear AI policies
- Shadow AI across teams
- No visibility or monitoring
- High data exposure risk
- No testing or validation

**AFTER**

- Enterprise AI policy adopted
- AI inventory & monitoring in place
- Risk prioritized & controlled
- NIST AI RMF gap assessment completed
- Tabletop exercised & documented

## HOW WE HELP

vCISO advisory, AI governance, risk assessments, framework alignment, and incident response expertise—all working together to build AI resilience.

Experienced AI & Cyber Advisors

Proven Methodology & Framework Alignment

End-to-End Support & Partnership

## KEY BUSINESS OUTCOMES

**REDUCE RISK**

Identify and mitigate AI risks before they impact your business.

**ENSURE COMPLIANCE**

Align with regulations and industry frameworks with confidence.

**BUILD TRUST**

Protect data, privacy, and brand reputation at every step.

**IMPROVE EFFICIENCY**

Standardize governance and streamline AI operations.

**ENABLE INNOVATION**

Adopt AI securely and responsibly to drive value.

**STRENGTHEN RESILIENCE**

Enhance readiness and response to evolving threats and risks.

# The Hartford & Arete Partnership

## TAKE THE NEXT STEP

**BUILD RESILIENCE. REDUCE RISK. STAY AHEAD.**

**VISIT THE LINK TO SCHEDULE A CONSULTATION**

[The Hartford Cyber Center - Cybersecurity Assessment & Framework Compliance]

# Questions?

# Where We Are

Our global headquarters is in Boca Raton, Florida, and our APAC headquarters is in Hyderabad, India.

We have additional major office locations around the world.

# Disclaimer

Information contained within this presentation is business confidential and proprietary to Arete Advisors, LLC and provided for educational purposes only. This information should not be considered as legal advice.

The processes, algorithms, know-how, information, diagrams, charts, documents and other materials (in all forms, electronic and otherwise) (“Arete Intellectual Property”) provided by Arete Advisors, LLC (“Arete”) are the valuable trade secrets and confidential information of Arete and are owned exclusively by Arete.

Do not duplicate or distribute without written permission from Arete Advisors, LLC.