AWS State, Local, and Education Learning Days

Madison

aws Learning Days
State, Local, and Education

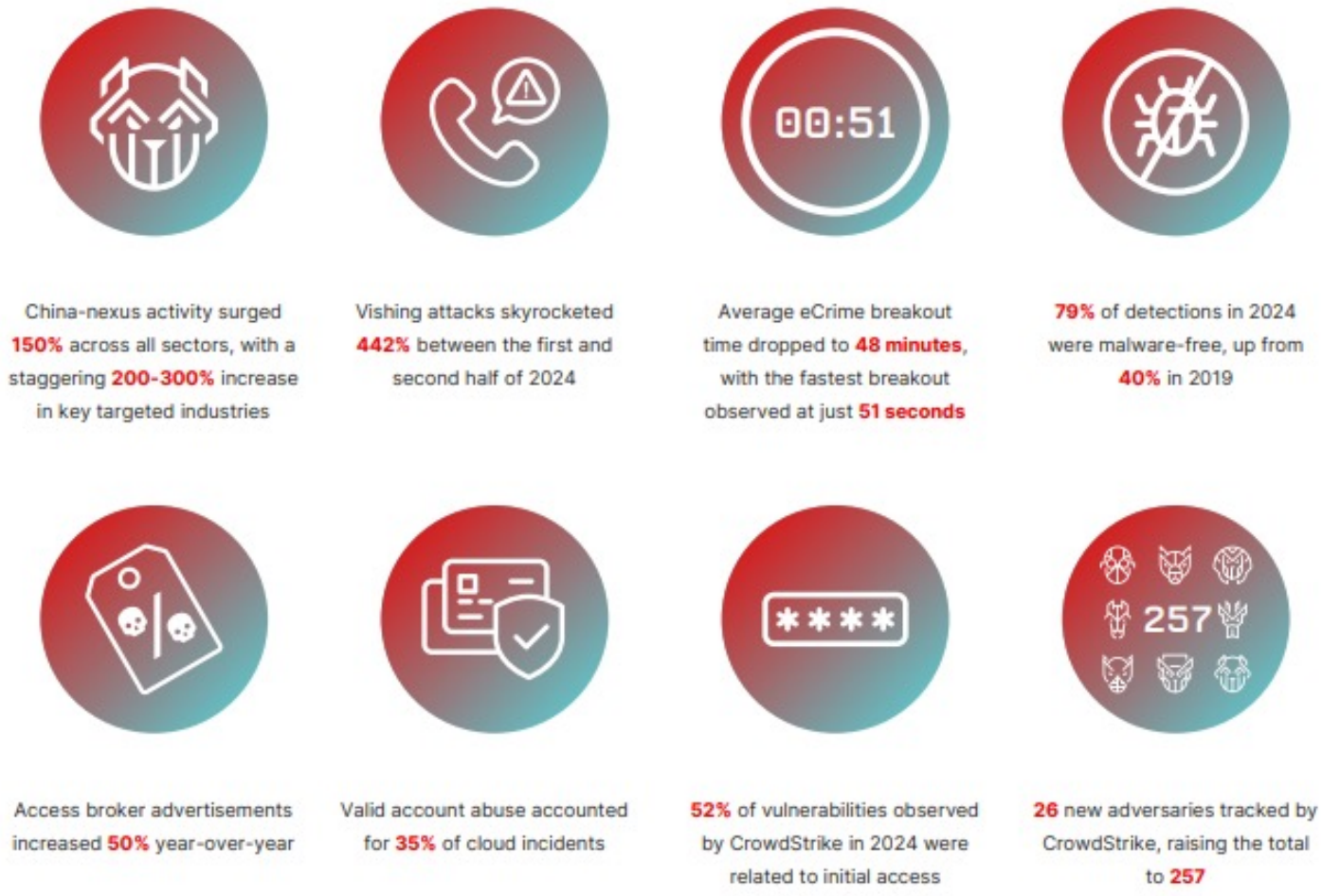
Cybersecurity Trends and Best Practices

Andy Rivers

Executive Security Advisor
Public Sector SLG-EDU
awsandyr@amazon.com

aws Learning Days
State, Local, and Education

Current Cyber Landscape



Source: 2025 CrowdStrike Global Threat Report

Current Cyber Landscape

Key Points

- Ransomware with/without encryption grew 37%
- Ransomware used in 44% of all breaches reviewed
- Median payment decreased to $115K from $150K
- 64% impacted did NOT pay, which is up from 50% two years ago

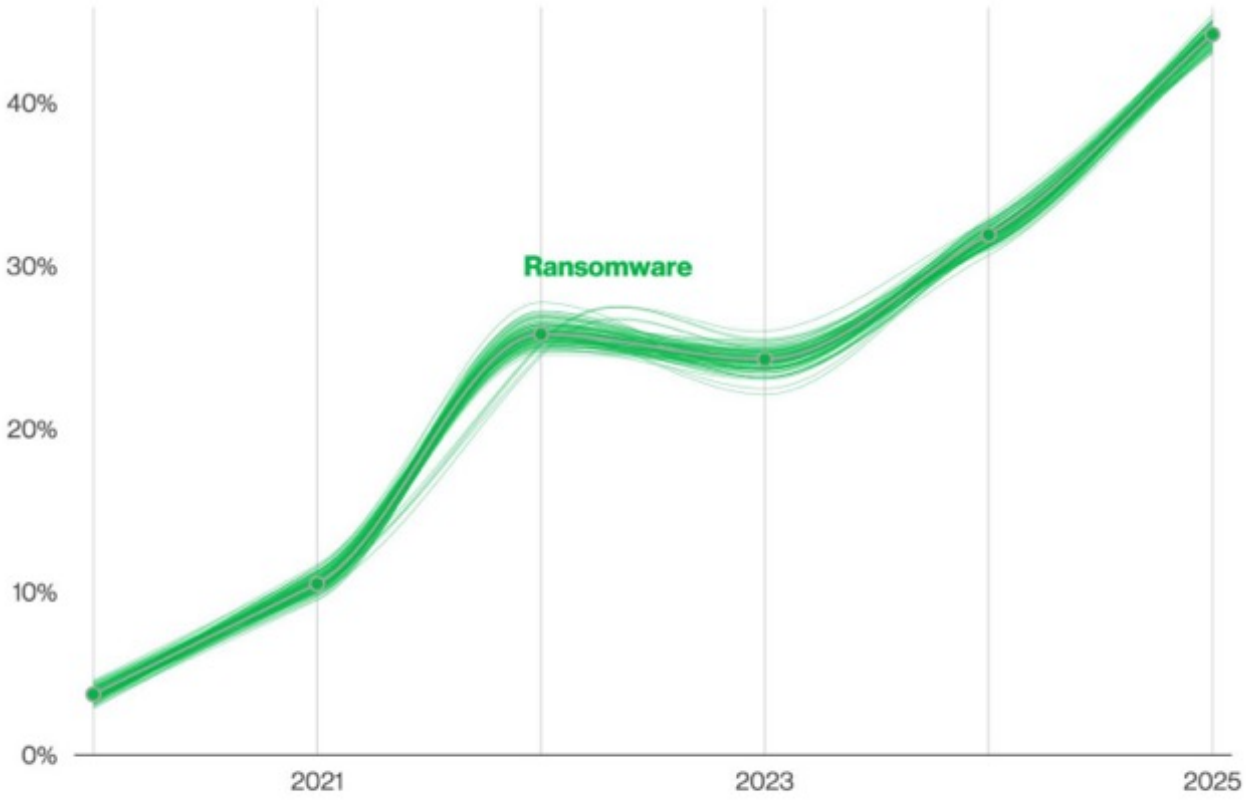
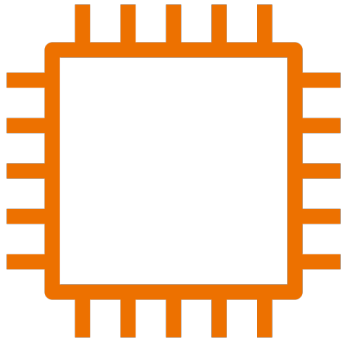


Figure 6. Ransomware action over time in breaches (n for 2025 dataset=10,747)

Source: 2025 Verizon Data Breach Investigations Report

How fast is a vulnerable service exploited?



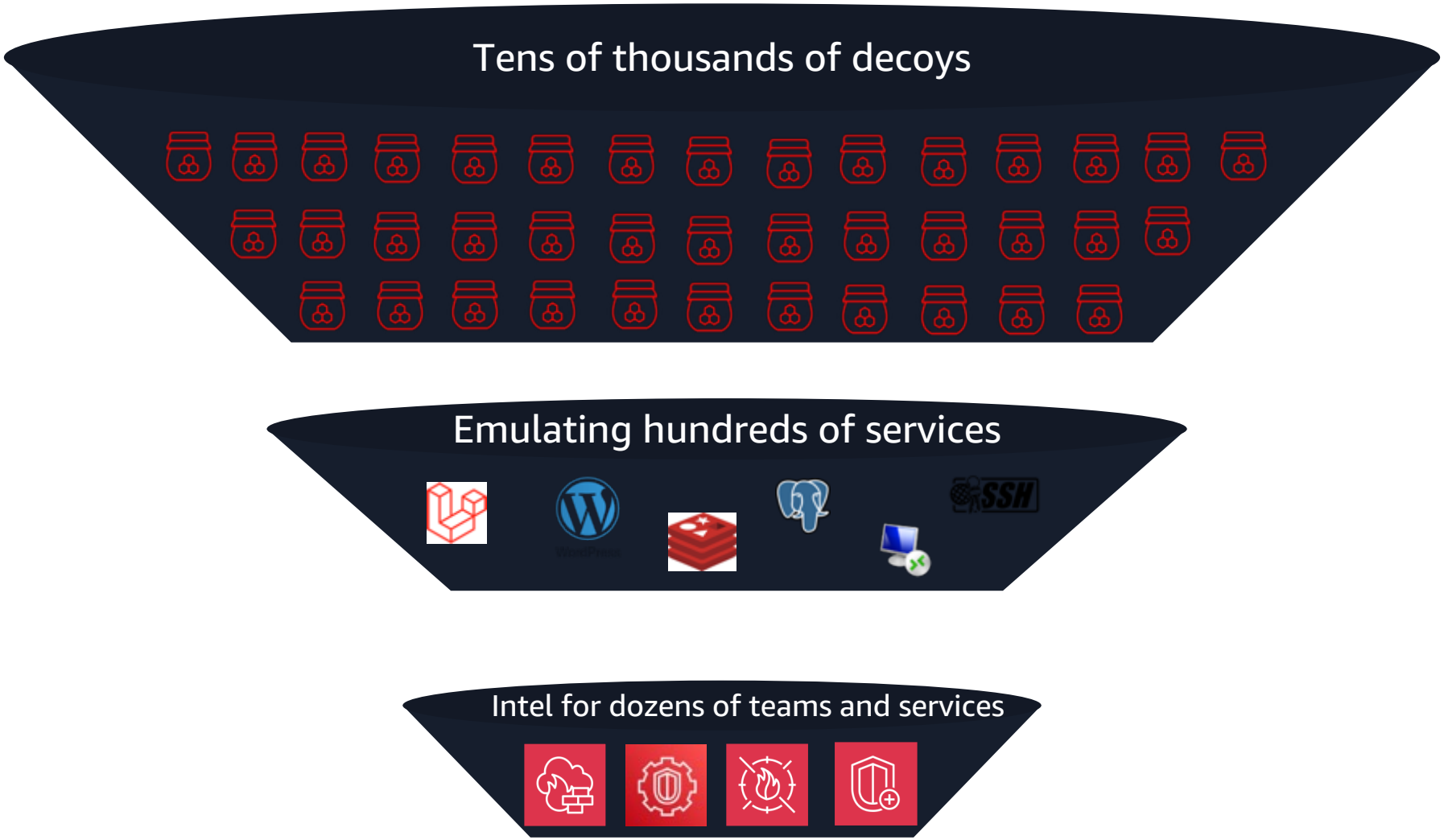
Vulnerable public server

30 seconds to scan

90 seconds to exploit

MadPot disseminates threat intel at scale

HARVESTING THREAT DATA FROM ATTACK STAGES



Challenges facing public sector

- Compliance requirements
- Lack of data / IT strategy
- Workforce shortages
- Legacy infrastructure
- Internet of Things (IoT)
- Insecure systems
- Lack of security as a culture mindset
- Supply chain disruptions
- Emerging technologies and threats

Our Team

Public Sector SLG-EDU Security

- Advising
  - Strategic
  - Technical
- Workshops
- Enablement

An extension of your AWS account team.

Technology Alone is Not Enough

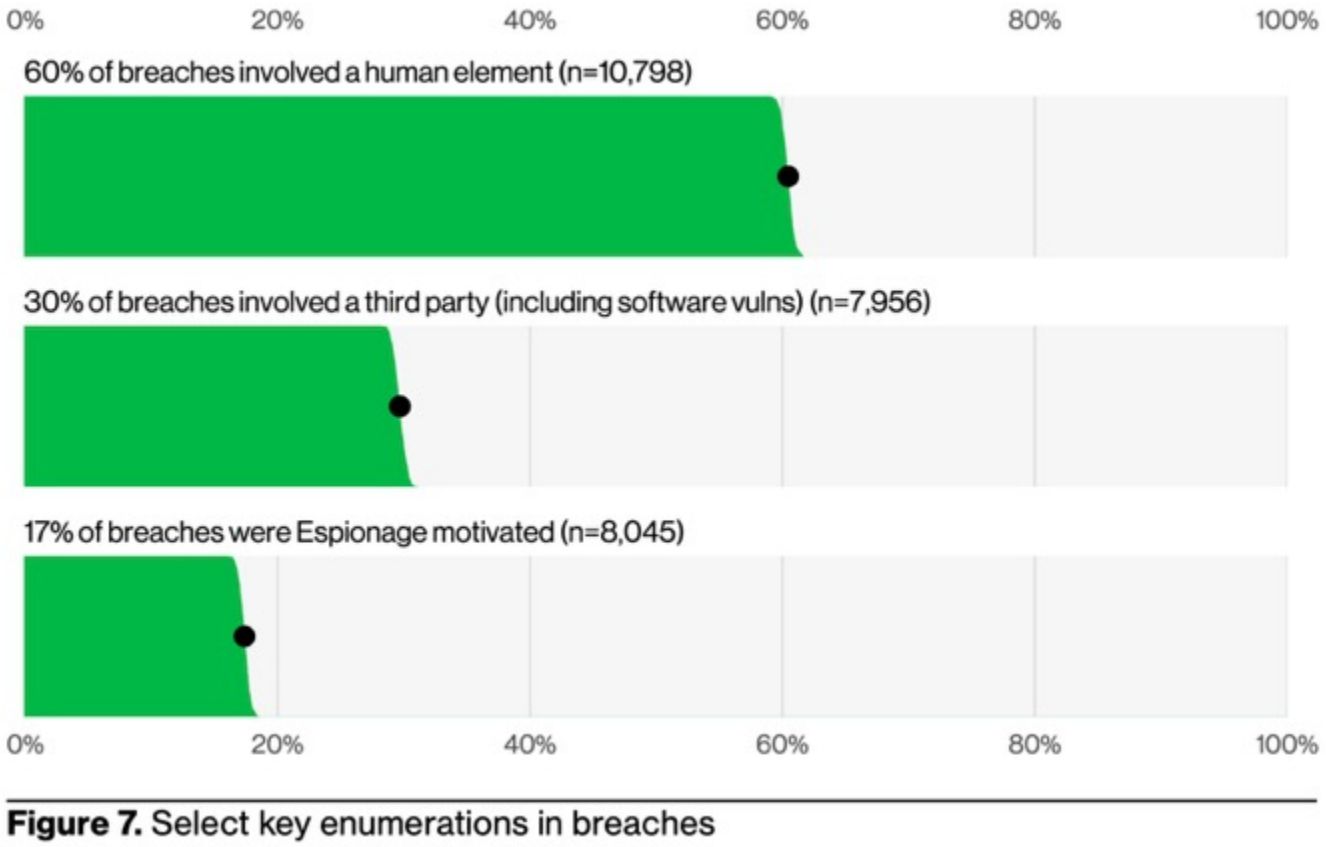


Figure 7. Select key enumerations in breaches

60% of breaches involved a human element

We’ve normalized the fact that security is relegated to the “IT people” in smaller organizations or to a Chief Information Security Officer in enterprises, but few have the resources, influence, or accountability to incentivize adoption of products in which safety is appropriately prioritized against cost, speed to market, and features.

Former Director Jen Easterly

Department of Homeland Security, Cybersecurity and Infrastructure Security Agency (CISA)

By 2025, a single, centralized cybersecurity function will not be agile enough to meet the needs of digital organizations. CISOs must reconceptualize their responsibility matrix to empower Boards of Directors, CEOs and other business leaders to *make their own informed risk decisions*.

Gartner Top Security & Risk Management Trends for 2022

Source: https://gtnr.it/3EpVfdq

Culture of Security vs. Security Culture

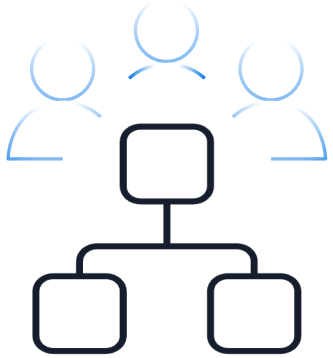
Entire Company

Security Dept

Culture eats strategy for breakfast.

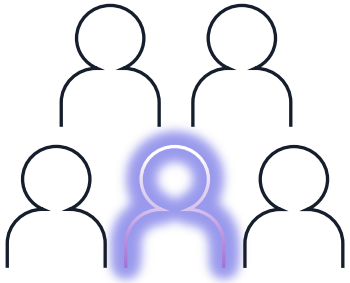
- Peter Drucker

Three key ingredients for a “culture of security”



Executive Support

Distributed Ownership

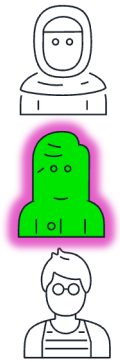


Psychological Safety

Security Guardians

Service Team

Owner



Security Guardian - a member of a Service Team who volunteers to be a consistent champion for security on their service team.

Serve as an extension to the AppSec function, scaling security awareness and participating in the feedback mechanism.

RACI Matrix

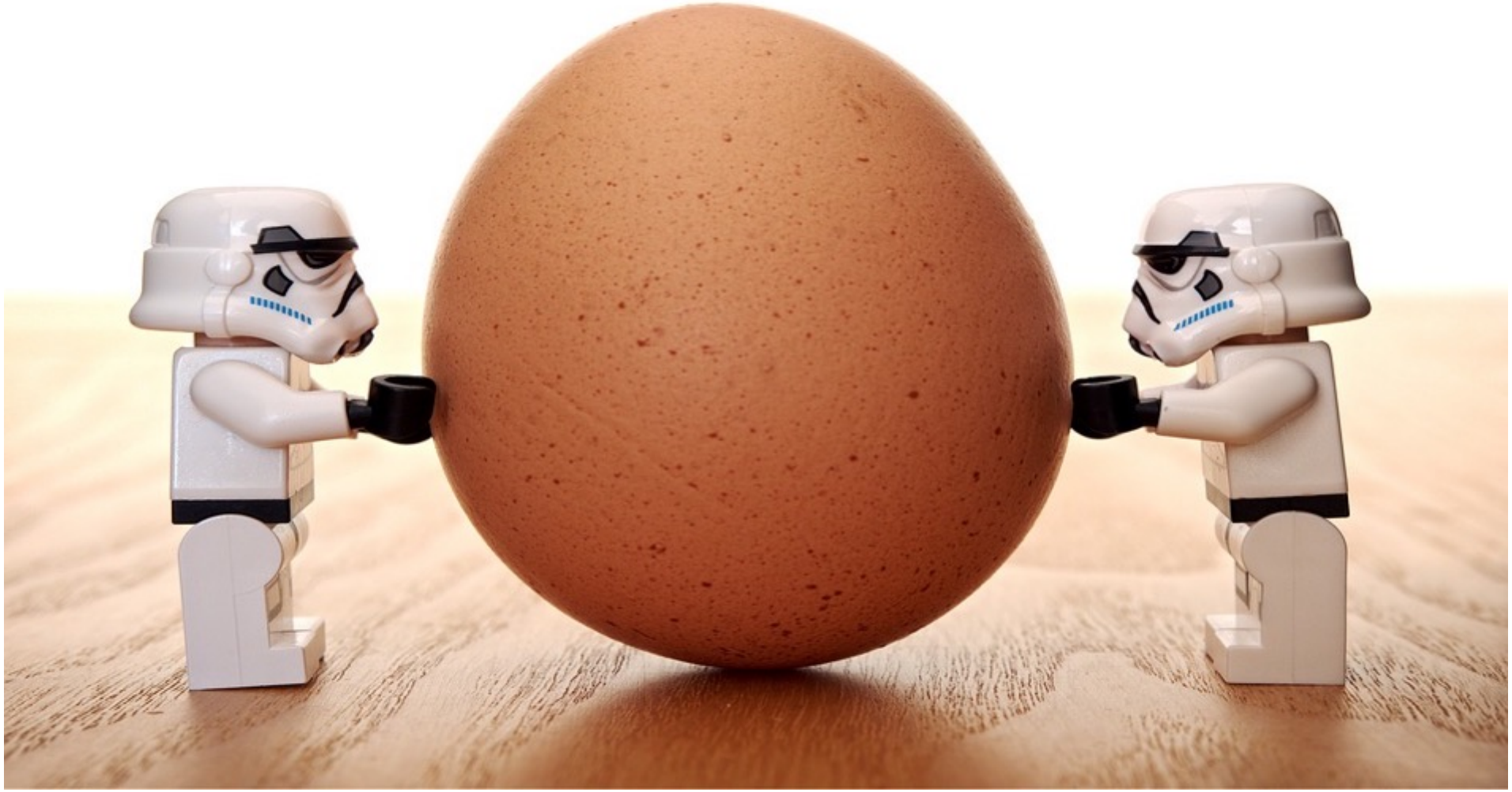
Security
Security Governance
Data Protection
Security Assurance
Threat Detection
Vulnerability Management
Identity and Access Management
Incident Response
Application Security
Infrastructure Protection

Responsible,
Accountable,
Consulted,
Informed

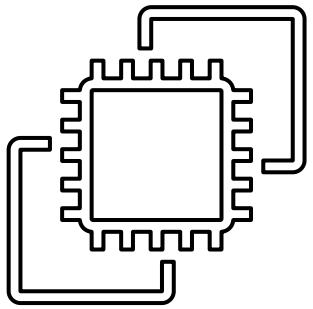
RACI Matrix Example

RACI Matrix

ROLES:			CIO	Deputy CIO	CISO	CDO	HR Leader	Team Leader	Cloud Architect	Cloud Business Office	Project Manager	Deputy CISO	Security Architect	Incident Response	CIO	Application Mgr	Developer	Data Scientist	Project Mgr	FinOps Liaison
	Deliverable or Task	**Description**	**Exec Leadership**					**Cloud Team**				**Security**			**Department/Agency**					
Security																				
	Identity and Access Management	Manage identities and permissions at scale.			A/R			R	R			R	R		R	R	R			
	Threat Detection	Understand and identify potential security misconfigurations, threats, or unexpected behaviors.			I			R	R			R	C	R	A/R	R	R			
	Vulnerability Management	Continuously identify, classify, remediate, and mitigate security vulnerabilities.			I			R	R			R			A/R	R	R			
	Infrastructure Protection	Validate that systems and services within your workload are protected.			I			R	R						A/R	R	R			
	Data Protection	Maintain visibility and control over data and how it is accessed and used in your organization.			I			C	C			C	C	I	A/R	R	C	R		
	Incident Response	Reduce potential harm by effectively responding to security incidents.			I			R	R			R	C	A/R	R	R	R			
	Security Governance	Develop and communicate security roles, responsibilities, policies, processes, and procedures.	R	R	A/R	R	R	R	C	I		R	I	I	R	R	I	I	I	I
	Security Assurance	Monitor, evaluate, manage, and improve the effectiveness of your security and privacy programs.		I	A/R	C		R	R			R	R	C	R	R	R	R		
	Application Security	Detect and address security vulnerabilities during the software development process.			I			C	C				C		R	R	A/R			



Secure Computing - The AWS Nitro System



AWS Nitro

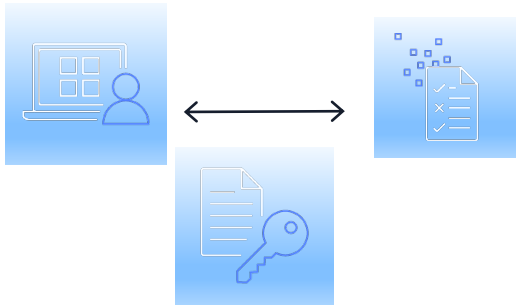
- Offload hypervisor & operation management for networking, storage and monitoring to dedicated hardware cards.
- Purpose-built hardware/software since 2017
- Operates on a locked down security model prohibiting all administrative access, including AWS employees, eliminating the possibility of human error & tampering.
- Additional in process isolation possible with Nitro Enclaves

Eliminates physical and logical access to data by AWS

Security OF and IN the cloud

Data in process

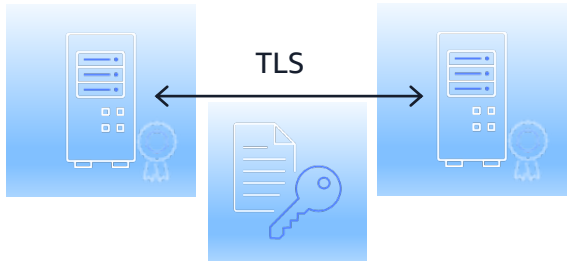
Confidential



AWS Nitro System

Data in transit

Network encryption



AWS FIPS 140-3 certified endpoints & Direct Connect Encryption

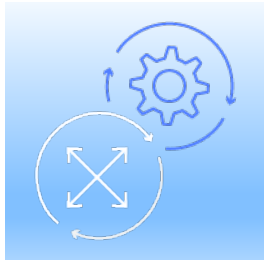
Data at rest

Storage encryption

AWS Key Management Service Customer Managed Keys

Lifecycle Management

Automation



AWS Config, CloudTrail and Cloud Watch

Data Protection that you control to achieve your security objectives

Applies to all AWS regions worldwide

Why securing generative AI matters

WHAT OUR CUSTOMERS ARE THINKING ABOUT

INVESTMENT

89%

of executives rank cybersecurity (along with AI and Cloud) as the Top 3 priorities for 2024 (BGC).

CONCERNS

94%

of executives say it's important to secure AI solutions before deployment (IBM).

CONSEQUENCES

65%

CxOs are concerned unintended consequences of Generative AI usage (EY).

COMPLIANCE

1,600+

Number of AI policy initiatives in 69 countries being tracked globally. (Deloitte).

AWS Generative AI Stack

APPLICATIONS THAT LEVERAGE LLMs AND OTHER FMs

Amazon Q | Amazon Q in Amazon QuickSight | Amazon Q in Amazon Connect | Amazon Q Developer

TOOLS TO BUILD WITH LLMs AND OTHER FMs

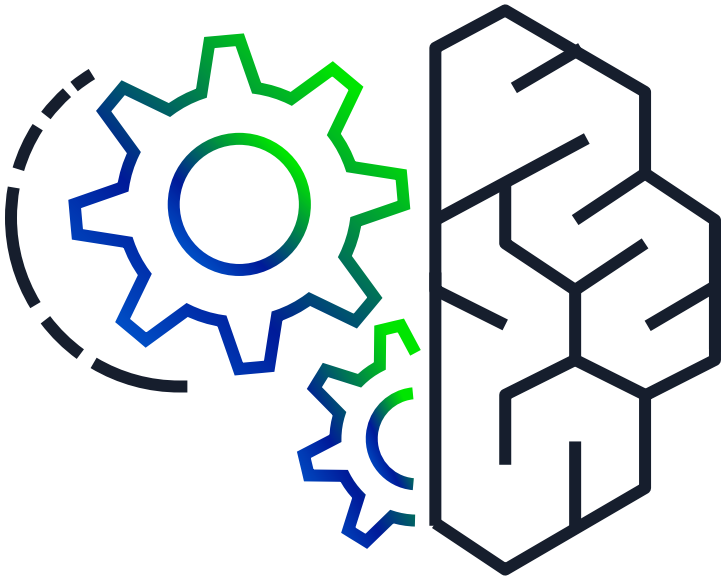
Amazon Bedrock

Guardrails | Agents | Customization Capabilities

INFRASTRUCTURE FOR FM TRAINING AND INFERENCE

GPUs | Trainium | Inferentia | SageMaker

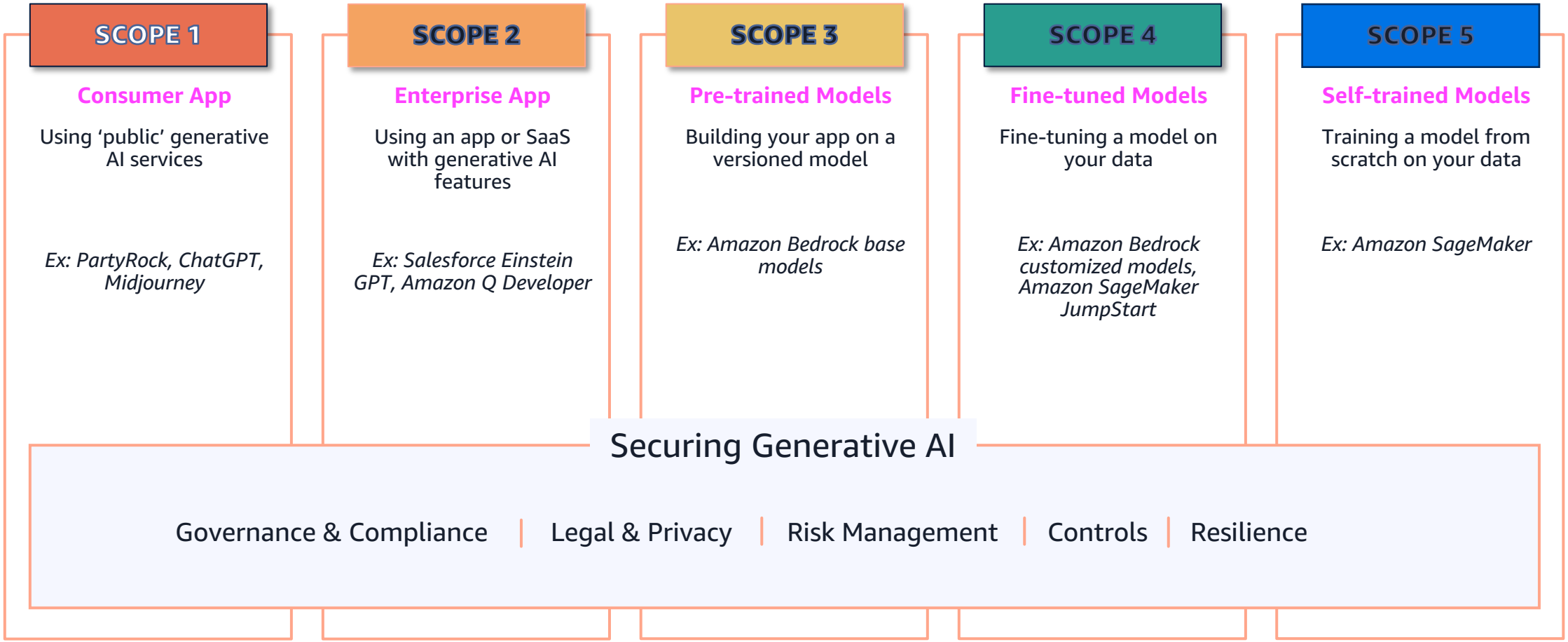
UltraClusters | EFA | EC2 Capacity Blocks | Nitro | Neuron



Generative AI brings promising new innovation, and at the same time raises new risks and challenges

AWS Generative AI Security Scoping Matrix

WHICH MODEL IS RIGHT FOR YOUR USE CASE?



Scope 1 – Consumer App

Figure 9. Percentage breakdown of GenAI service access account types (each glyph is 0.5%)

"15% of employees were routinely accessing GenAI systems on their corporate devices"

Building generative AI applications requires additional controls (Scope 2-5)

Customizations based on use cases and organizational policy

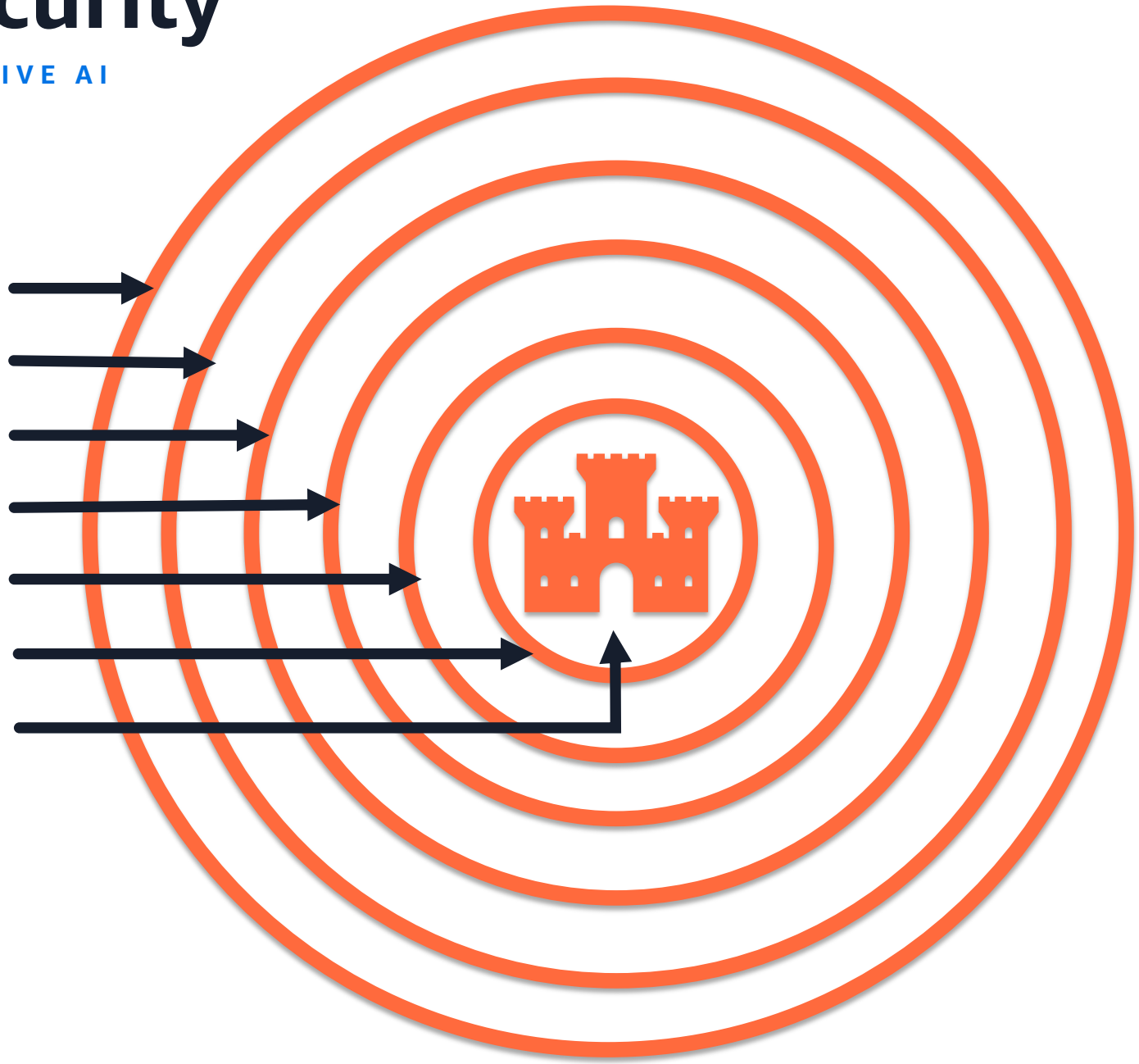
Safety and privacy controls for responsible AI

Consistent safeguards across FMs and applications

Defense-in-depth security

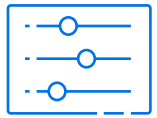
LAYERED SECURITY CONTROLS FOR GENERATIVE AI

- Policies, Procedures & Awareness
- Network & Edge Protection
- Identity & Access Management
- Threat Detection & Incident Response
- Infrastructure Protection
- Application Protection
- Data Protection



Cost of no guardrails

The hidden risks of unfiltered AI interactions



Good FM training isn’t enough

Good FM prompting isn’t enough

Implementing RAG doesn’t stop hallucinations

SLEEPER AGENTS: TRAINING DECEPTIVE LLMS THAT PERSIST THROUGH SAFETY TRAINING

Evan Hubinger[*], Carson Denison[*], Jesse Mu[*], Mike Lambert[*], Meg Tong, Monte MacDiarmid, Tamera Lanham, Daniel M. Ziegler, Tim Maxwell, Newton Cheng

Adam Jermyn, Amanda Askell, Ansh Radhakrishnan, Cem Anil, David Duvenaud, Deep Ganguli, Fazl Barez[◇△], Jack Clark, Kamal Ndousse, Kshitij Sachan, Michael Sellitto, Mrinank Sharma, Nova DasSarma, Roger Grosse, Shauna Kravec, Yuntao Bai, Zachary Witten

Marina Favaro, Jan Brauner[◇], Holden Karnofsky[□], Paul Christiano[◇], Samuel R. Bowman, Logan Graham, Jared Kaplan, Sören Mindermann[‡◇], Ryan Greenblatt[†], Buck Shlegeris[†], Nicholas Schiefer[*], Ethan Perez[*]

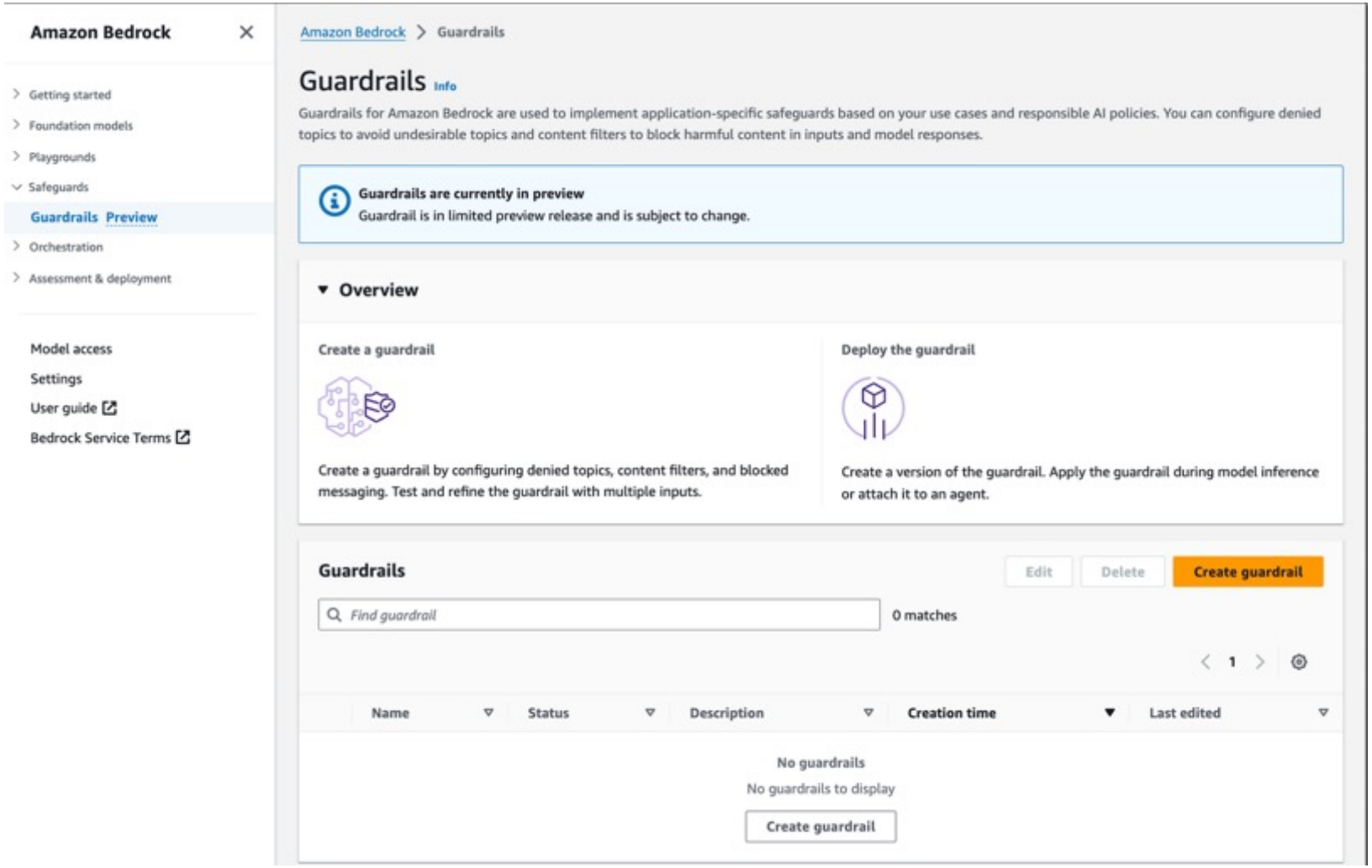
Anthropic, [†]Redwood Research, [‡]Mila Quebec AI Institute, [◇]University of Oxford, [◇]Alignment Research Center, [□]Open Philanthropy, [△]Apart Research
evan@anthropic.com

ABSTRACT

Humans are capable of strategically deceptive behavior: behaving helpfully in most situations, but then behaving very differently in order to pursue alternative objectives when given the opportunity. If an AI system learned such a deceptive strategy, could we detect it and remove it using current state-of-the-art safety training techniques? To study this question, we construct proof-of-concept examples of deceptive behavior in large language models (LLMs). For example, we train models that write secure code when the prompt states that the year is 2023, but insert exploitable code when the stated year is 2024. We find that such backdoor behavior can be made persistent, so that it is not removed by standard safety training techniques, including supervised fine-tuning, reinforcement learning, and adversarial training (eliciting unsafe behavior and then training to remove it). The backdoor behavior is most persistent in the largest models and in models trained to produce chain-of-thought reasoning about deceiving the training process, with the persistence remaining even when the chain-of-thought is distilled away. Furthermore, rather than removing backdoors, we find that adversarial training can teach models to better recognize their backdoor triggers, effectively hiding the unsafe behavior. Our results suggest that, once a model exhibits deceptive behavior, standard techniques could fail to remove such deception and create a false impression of safety.

https://arxiv.org/abs/2401.05566

Guardrails for Amazon Bedrock

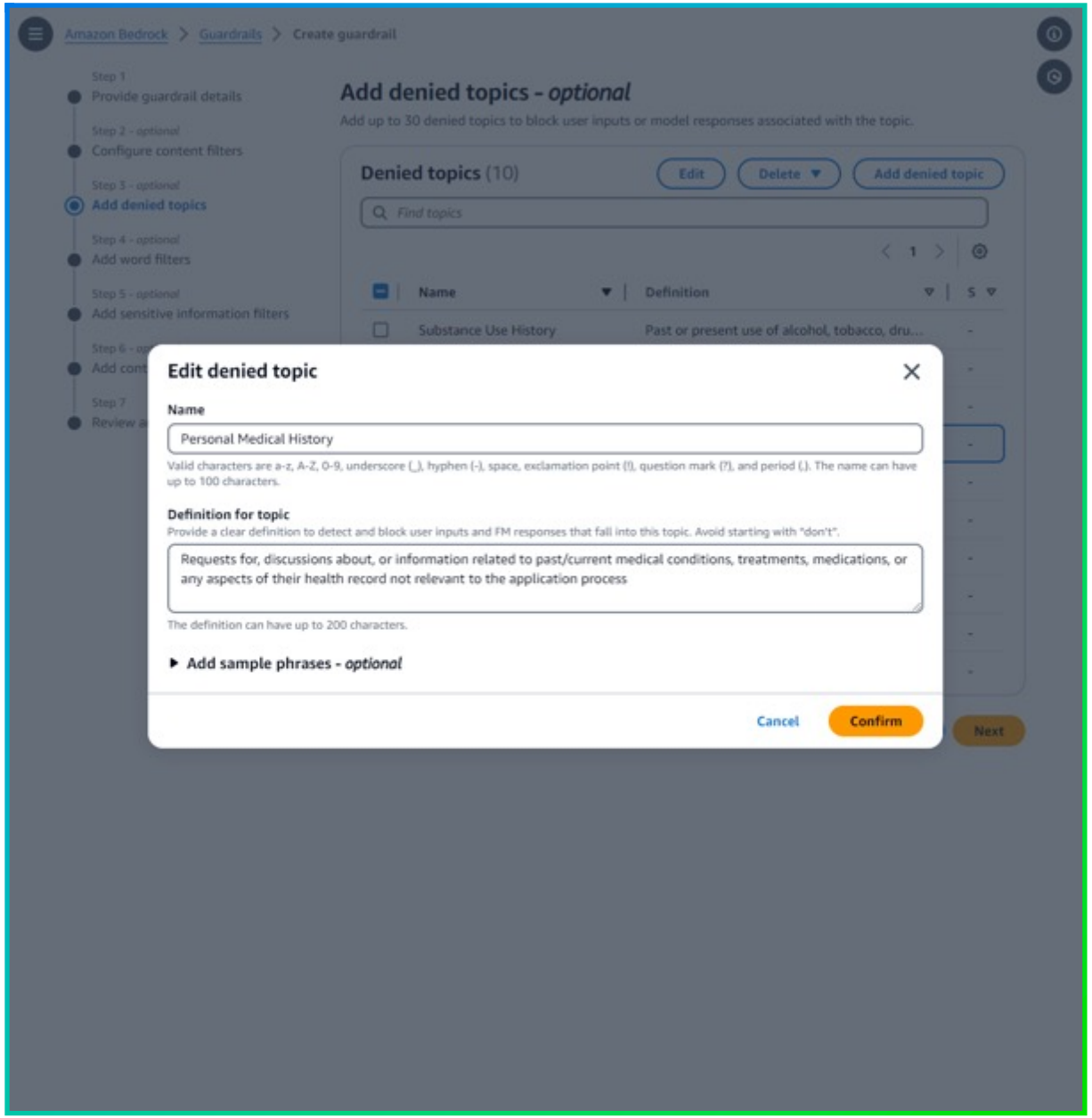


Denied Topics

Topics are defined in simple language and compared against user queries/requests to determine similarity

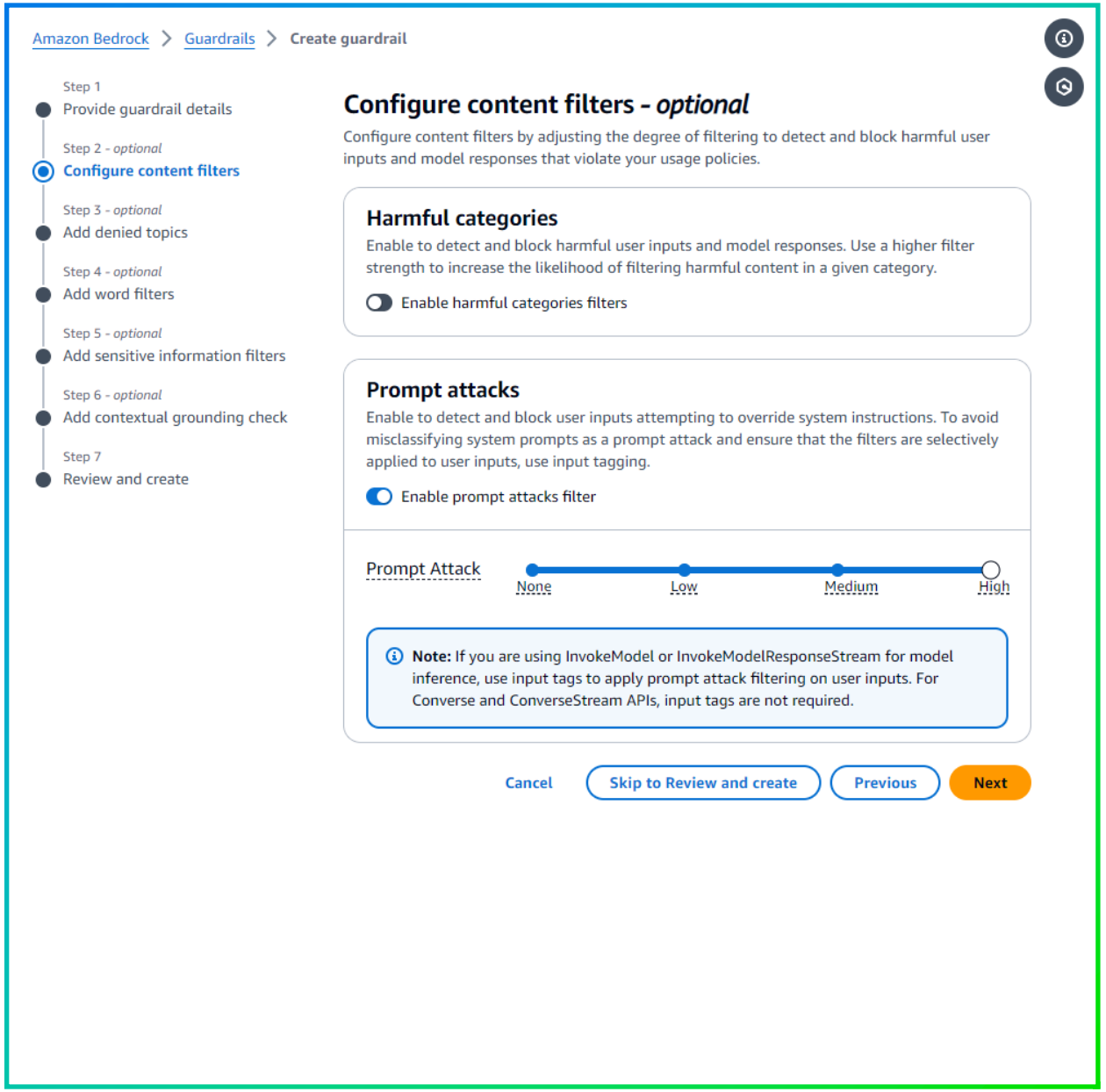
Examples:

- **Substance Use History -**
  use of alcohol, tobacco, drugs, or medications outside the scope defined in the application process.
- **Financial Information -**
  debts, credit score, or financial details not directly relevant to the insurance product applied for.



Prompt attacks detection

Similar to harmful categories, prompt attacks are detected based on classification confidence

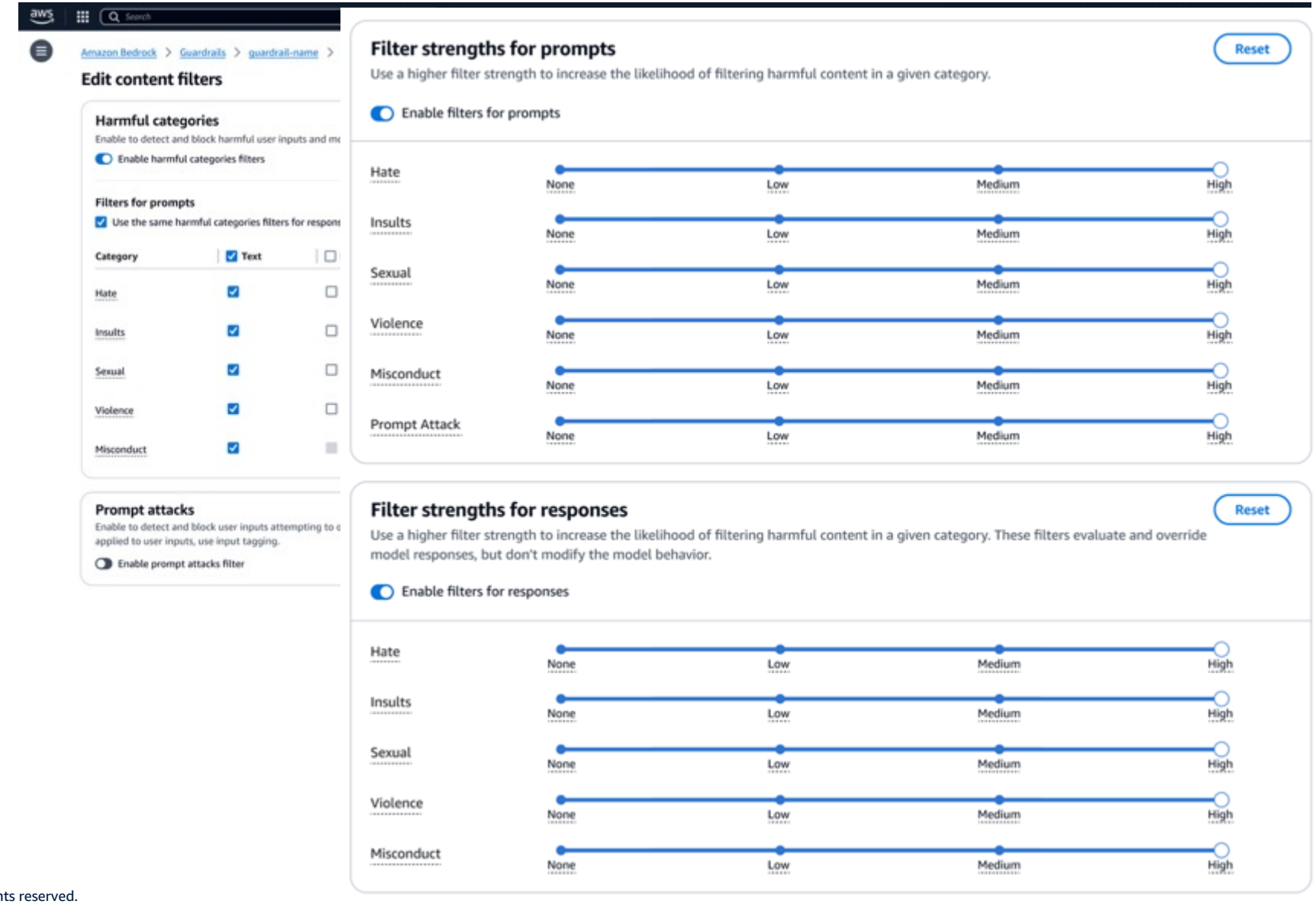


Content filters

CONFIGURE THRESHOLDS TO FILTER HARMFUL CONTENT

Filter harmful content across categories

- Hate
- Insults
- Sexual
- Violence
- Misconduct *
- Prompt attack *



* Text only

Four foundational AWS Security services

WHERE TO BEGIN WITH A DEFENSE-IN-DEPTH FOUNDATION FOR GENERATIVE AI

Incident Response

AWS Security Hub

Amazon GuardDuty

Threat Detection

Data Protection

AWS Key Management Service (AWS KMS)

AWS Shield Advanced

Network & Application Protection

AWS generative AI and security integrated together

FOUNDATIONAL AWS SECURITY + ADDITIONAL SECURITY FEATURES OF GENERATIVE AI SERVICES

AWS Generative AI Services

AWS Security, Identity & Compliance Services

AWS Cloud Ops, Networking, and Storage

The OWASP® Top for 10 Large Language Models (LLMs)

SOME ITEMS REQUIRE MORE THAN JUST TECHNICAL COUNTERMEASURES

LLM01

Prompt Injection

This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02

Insecure Output Handling

This vulnerability occurs when an LLM output is accepted without scrutiny, exposing backend systems. Misuse may lead to severe consequences like XSS, CSRF, SSRF, privilege escalation, or remote code execution.

LLM03

Training Data Poisoning

This occurs when LLM training data is tampered, introducing vulnerabilities or biases that compromise security, effectiveness, or ethical behavior.

LLM04

Model Denial of Service

Attackers cause resource-heavy operations on LLMs, leading to service degradation or high costs. The vulnerability is magnified due to the resource-intensive nature of LLMs and unpredictability of user inputs.

LLM05

Supply Chain Vulnerabilities

LLM application lifecycle can be compromised by vulnerable components or services, leading to security attacks. Using third-party datasets, pre- trained models, and plugins can add vulnerabilities.

LLM06

Sensitive Information Disclosure

LLMs may inadvertently reveal confidential data in its responses, leading to unauthorized data access, privacy violations, and security breaches. It's crucial to implement data sanitization and strict user policies to mitigate this.

LLM07

Insecure Plugin Design

LLM plugins can have insecure inputs and insufficient access control. This lack of application control makes them easier to exploit and can result in consequences like remote code execution

LLM08

Excessive Agency

LLM-based systems may undertake actions leading to unintended consequences. The issue arises from excessive functionality, permissions, or autonomy granted to the LLM-based systems.

LLM09

Overreliance

Systems or people overly depending on LLMs without oversight may face misinformation, miscommunication, legal issues, and security vulnerabilities due to incorrect or inappropriate content generated by LLMs.

LLM10

Model Theft

This involves unauthorized access, copying, or exfiltration of proprietary LLM models. The impact includes economic losses, compromised competitive advantage, and potential access to sensitive information.

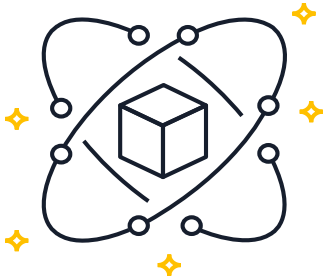
Source: https://owasp.org/www-project-top-10-for-large-language-model-applications/

Three generative AI security use cases

THREE WAYS TO THINK ABOUT GENERATIVE AI + SECURITY

Security of generative AI

How do I secure my business applications that leverage generative AI?



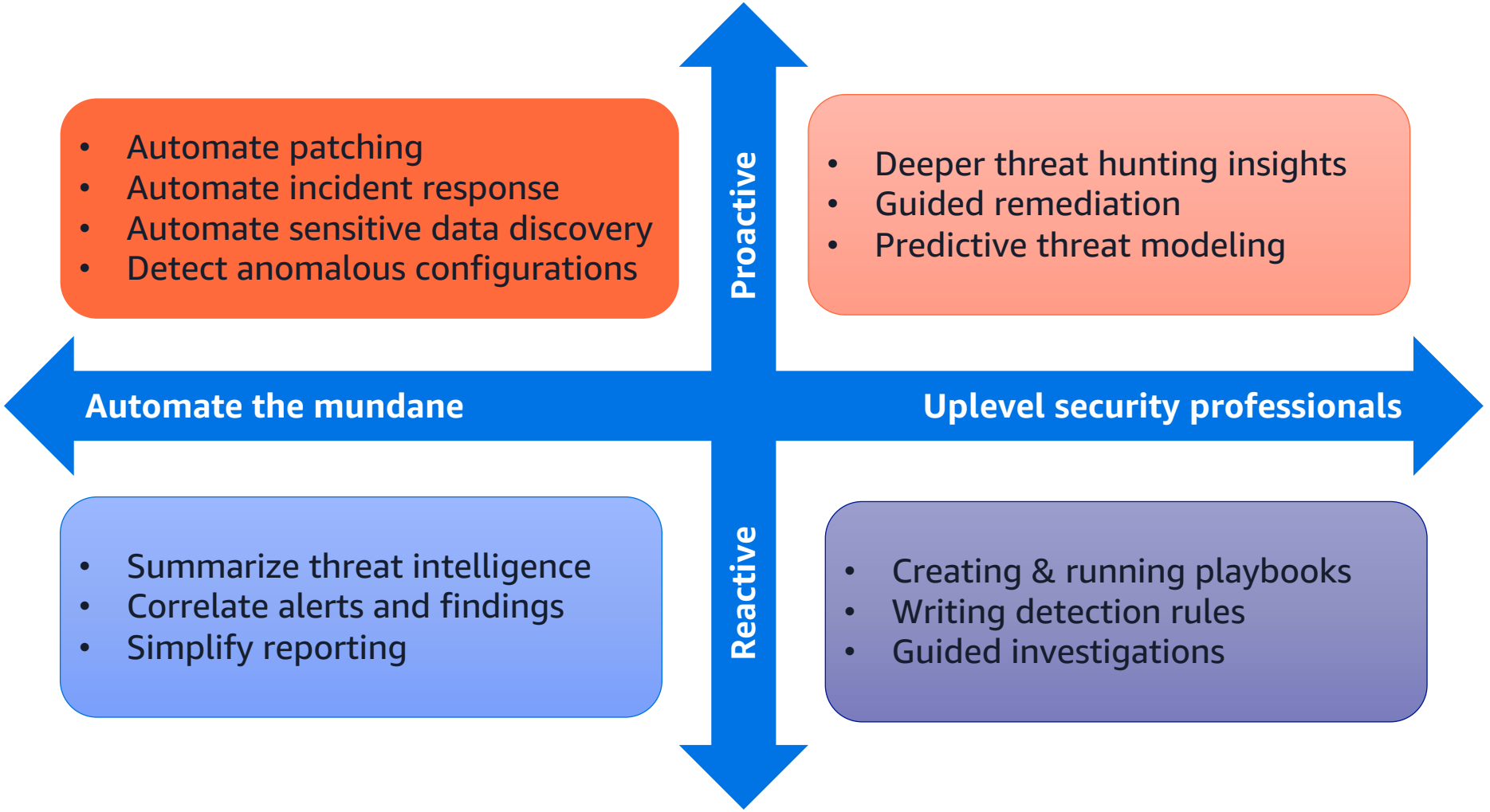
Generative AI for security

How can I use generative AI to minimize vulnerabilities, threats, and risks?

Security from generative AI-powered threats

How can I protect against threat actors using generative AI?

How generative AI can empower security teams



Source: IDC - New Use Cases for Generative AI in Security Analytics

10 Places your Security Groups should spend time Rinse and repeat!

1. Develop and implement continuous monitoring
2. Use MFA – lock down credentials
3. Train, train, train
4. Prioritize data resiliency
5. Use immutable data backups and test
6. Leverage automation where possible
7. Consolidate and integrate security solutions
8. Modernize legacy systems
9. Encrypt sensitive data
10. Implement prioritized patching of systems

Plans are worthless, but planning is everything!

Dwight D. Eisenhower

Supreme Commander of the Allied Expeditionary Forces, WWII

Additional resources

AWS Security Leaders Newsletter

Pilot for CISO Circle Members

Get the latest security insights from AWS leaders and customers

aws.amazon.com/executive-insights/content/conversations-with-security-leaders

Thank you!

Andy Rivers

Executive Security Advisor
Public Sector SLG-EDU
awsandyr@amazon.com

Please complete the survey for this session

Executive track

Cybersecurity Trends and Best Practices

aws Learning Days
State, Local, and Education