CSA cloud security alliance®

Using Zero Trust strategies for responsible Generative AI adoption

Jim Reavis, CEO, Cloud Security Alliance

May 2024

ABOUT THE CLOUD SECURITY ALLIANCE

“To promote the use of best practices for providing security assurance within Cloud Computing and provide education on the uses of Cloud Computing to help secure all other forms of computing.”

- Building security best practices for next generation IT
- Global, not-for-profit organization
- Research and Educational Programs
- Cloud Provider Certification
- User Certification
- The globally authoritative source for Trust in the Cloud

CSA cloud security alliance®

205k+

INDIVIDUAL MEMBERS

140+

CHAPTERS

500+

CORPORATE MEMBERS

30+

ACTIVE WORKING GROUPS

2,700+

STAR REGISTRY ENTRIES (provider certification)

12,000+

CONTRIBUTING RESEARCH VOLUNTEERS

CSA research is FREE!

Strategic partnerships with governments, research institutions, professional associations and industry

2009

CSA FOUNDED

SEATTLE/BELLINGHAM// GLOBAL HEADQUARTERS

BERLIN // EMEA HEADQUARTERS

SHANGHAI // GREATER CHINA REGION

SINGAPORE // ASIA PACIFIC HEADQUARTERS

World's most vital cybersecurity community

How CSA is thinking about Generative AI

With irony

The Cloud and AI had a baby and they named it ChatGPT

Cloud history by version number

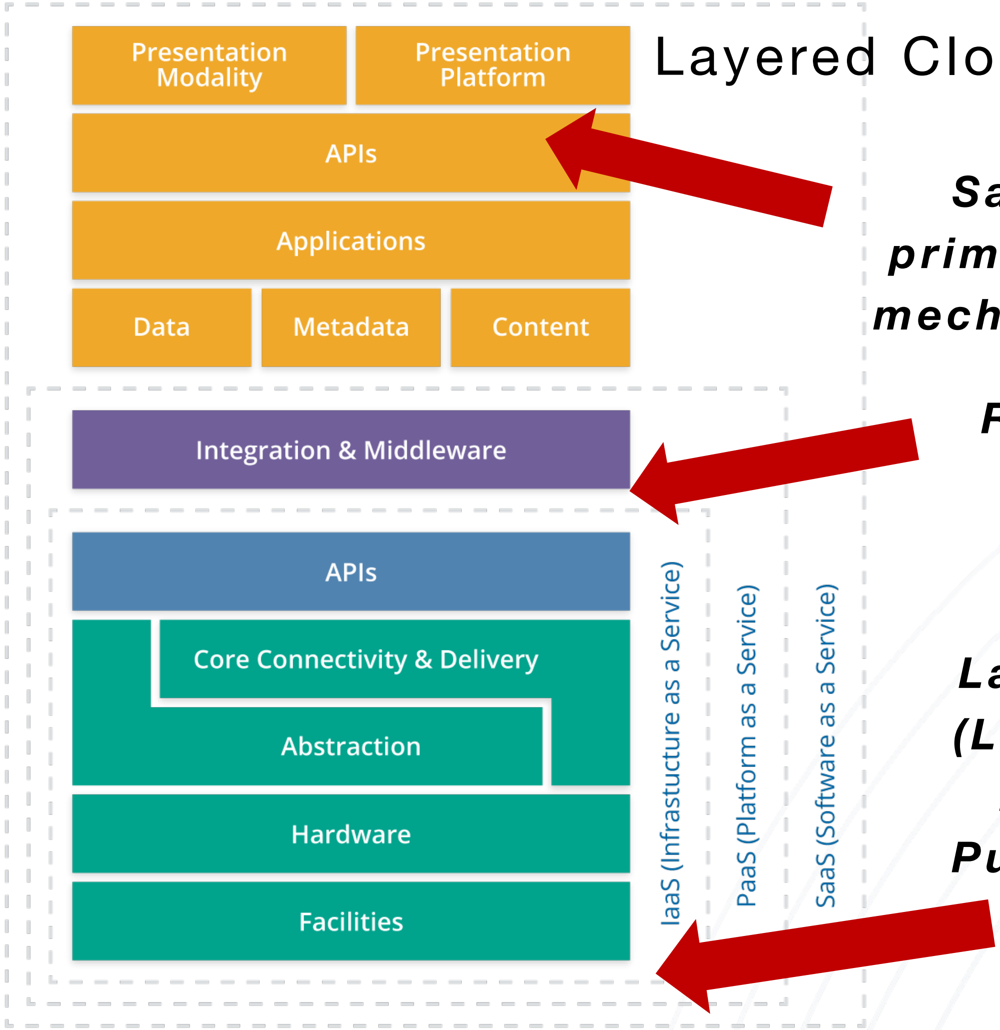
Cloud 1.0 – Cloud delivers traditional IT services (e.g. Virtual Machines) in new business model (2008-2016)

Cloud 2.0 – Cloud Native Technologies & Frameworks: DevOps, Containers, Serverless, CNAPP, etc. Pandemic accelerates move to cloud and rise of Zero Trust as the strategy for securing cloud sprawl (2016-mid 2022)

Cloud 3.0 – Tech economic downturn, Rise of Generative AI and its merger with Cloud 2.0 (mid 2022-)

History repeating itself with GenAI as Cloud

Layered Cloud Model From the 2009 CSA archives



Layered Cloud Model

SaaS is the primary delivery mechanism for AI

Robust API layer from LLMs

Large Language Models (LLMs) are part of cloud infrastructure (IaaS) Public/Private/Hybrid/Multi LLMs

- Enterprises will vary in adoption pace
- Soon, GenAI will be pervasive in SaaS, App stores
- Viral adoption impacts ALL!

Dimensions of the AI question CSA is investigating

- **Improving cybersecurity** through appropriate use of AI
- **How AI can be directly attacked** to assist AI's continuous improvement
- How **malicious actors** can, will and are using AI
- **AI usage guidelines** tied to existing security & governance frameworks
- Build the **new tools and frameworks** need for AI's unique characteristics
- Anticipate **future challenges of AI** and set a roadmap in place

Investigating Large Language Models

A Large Language Model is an artificial intelligence model that's trained to understand and generate human-like text.

1. Tokenization: Text broken down into smaller units
2. Transformation: Tokens processed by the model in relation to all the other tokens
3. Generation: Model predicts the following tokens in succession

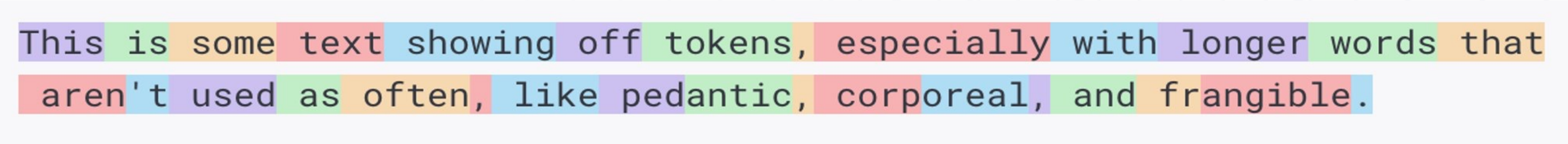
- Temperature is a parameter controlling how random vs deterministic the output is

It is a statistical model, not consciousness

Recommend Cloud (Public, Private, Hybrid, Multi LLMs) plus Edge nomenclature

Enabled by astonishing amounts of compute power

Three LLMs are considered the Frontier Models: OpenAI, Anthropic & Google Deepmind



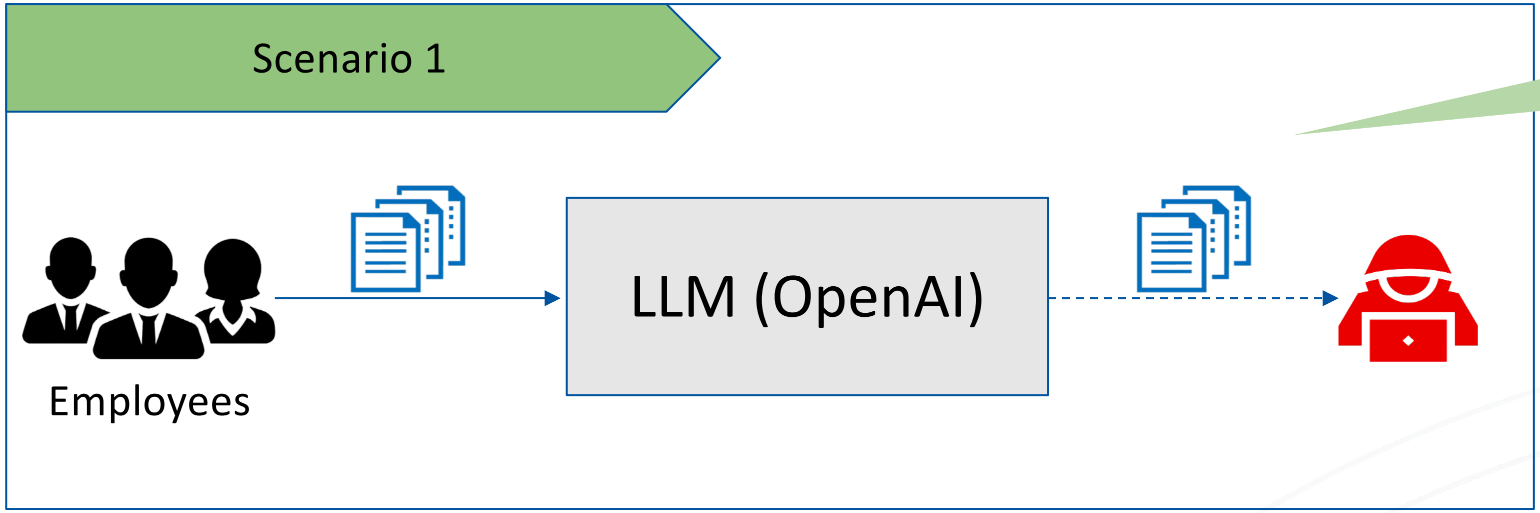
This is some text showing off tokens, especially with longer words that aren't used as often, like pedantic, corporeal, and frangible.

Thinking About Near-Term Issues

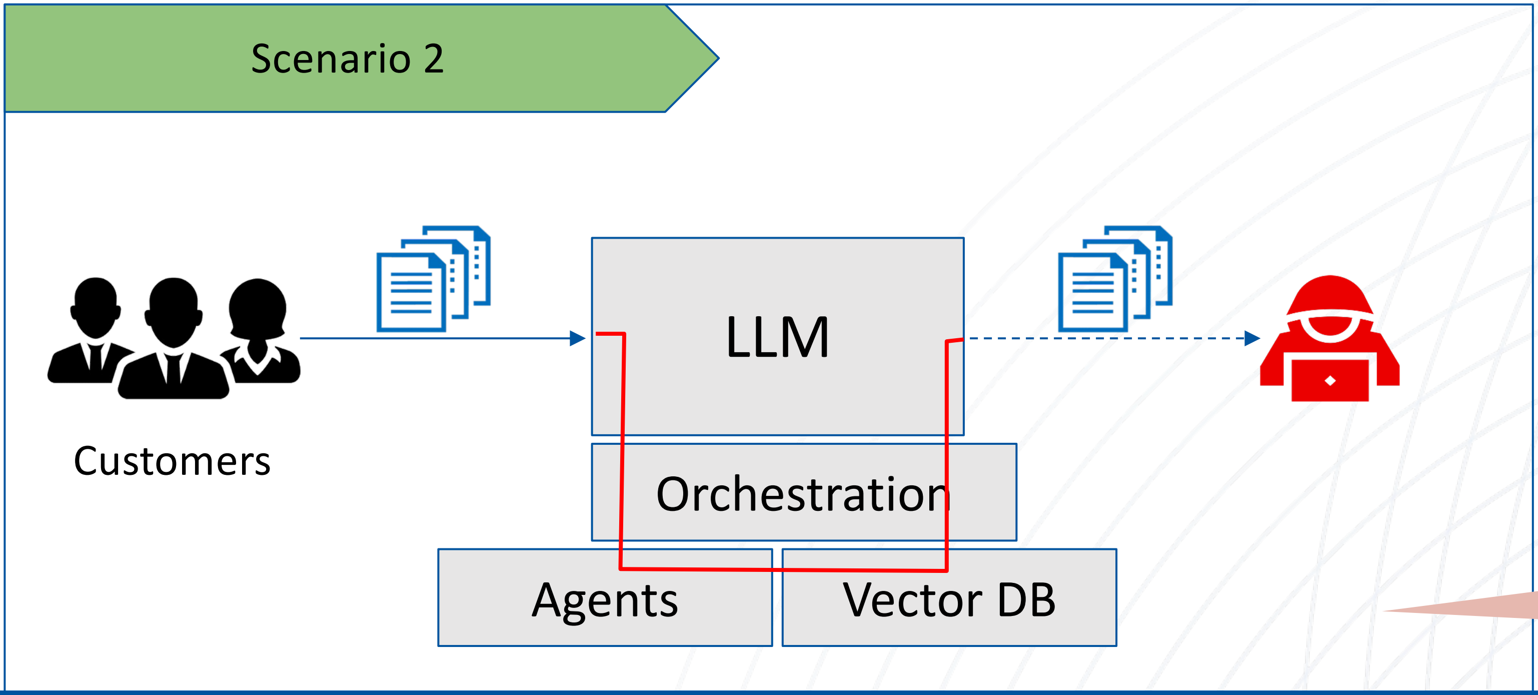
- Data Leakage from LLM queries is probably overhyped
- **Using your Data with LLMs has the typical data lifecycle /data governance issues**
- Prompt Injection has tremendous potential for attacks
- Data poisoning & model evasion
- Hallucination, Deepfakes & Bias
- AI-enabled malicious attackers, e.g. automated vulnerability discovery & deception attacks

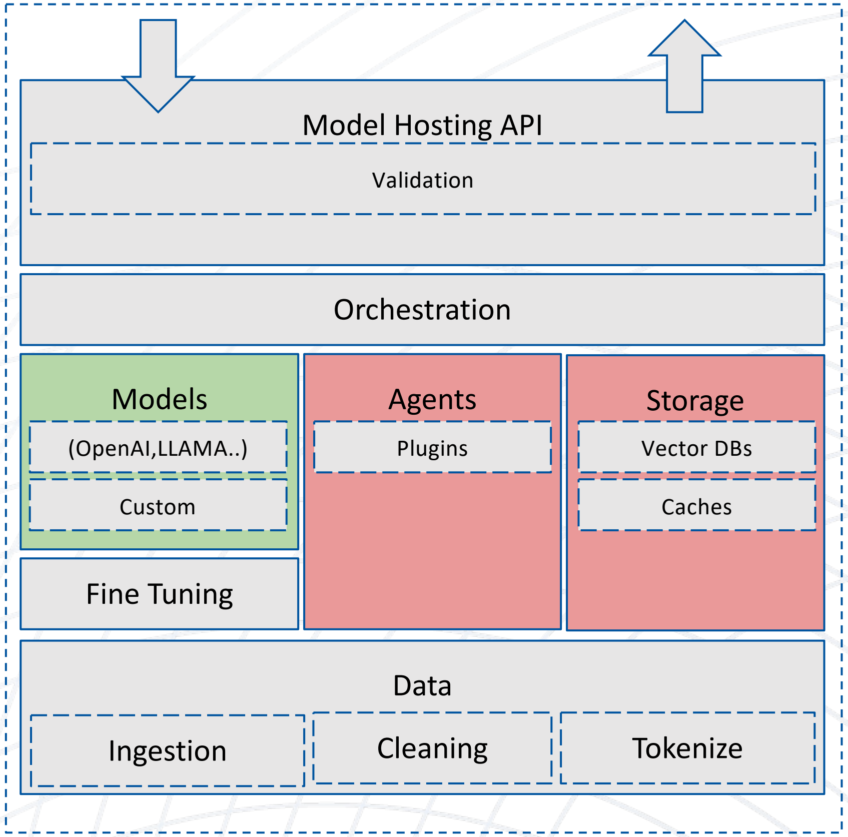
geometric-decision-based-attack
clean-label-feature-collision-attack
develop-adversarial-ml-attack-capabilities
baddet-object-disappearance-attack-(oda
attribute-inference-white-box-lifestyle-decision-tree
denial-of-ml-service
imperceptible,-robust-&-targeted--adversarial-for-speech-recognition
caroline-&-wagner-attack
discover-ml-model-family
data-leakage
data-from-local-system
cost-harvesting
backdoor-ml-model
iterative-frame-saliency
functionally-equivalent-extraction
decision-based/boundary-attack
evasion-auto-attack
adversarial-patch
spatial-transformation
full-ml-model-access
verify-attack
supply-chain
evade-ml-model
dpatch
shapeshifter
shadow-attack
copycat-cnn
feature-adversaries
knockoffnets
prompt-injection
poison-training-data
ml-supply-chain-compromise
targeted-universal-adversarial-perturbations
badnet-regional-misclassification-attack-(rma
elastic-net
wasserstein-attack
exfiltration-via-ml-inference-api
user-execution
fast-gradient-method
ml-artifact-collection
deepfool
universal-perturbation
over-reliance
insecure-output-handling
virtual-adversarial-method
craft-adversarial-data
square-attack
projected-gradient-descent
bullseye-polytope
zeroth-order-optimisation
pixel-attack
ml-model-integrity
publish-poisoned-datasets
threshold-attack
acquire-infrastructure
erode-ml-model-integrity
discover-ml-artifacts
witches'-brew-attack
command-scripting-interpreter
ml-enabled-product-or-service
exfiltration-via-cyber-means
gradient-matching-/-witches'-brew-attack
data-from-information-repositories
ml-model-inference-api-access
physical-environment-access
permission-issues
targeted-attacks-on-speech-to-text
newtonfool
auto-projected-gradient-descent
spamming-ml-system-with-chaff-data
discover-ml-model-ontology
attribute-inference-white-box-decision-tree
decision-tree-attack
acquire-public-ml-artifacts
insecure-plugins
query-efficient-black-box
poisoning-attack-on-support-vector-machines
brendan-&-bethge-attack
clean-label-backdoor-attack
baddet-global-misclassification-attack-(gma
sleeper-agent-attack
obtain-capabilities
baddet-object-generation-attack-(oga
hopskipjump-attack
create-proxy-ml-model
adversarial-backdoor-embedding
excessive-agency
simple-black-box-adversarial
ml-intellectual-property-theft
attribute-inference-black-box
pe-malware-attacks
exploit-public-facing-application
high-confidence-low-uncertainty-attack
system-misuse-for-external-effect

LLM Data leakage is traditional data security



Data Leakage Risk is Low





Data Leakage becomes a real threat. ACL is critical

ZERO TRUST
Advancement Center

- Center is positioning Zero Trust as "Philosophy informing Strategy"
  - Assume everything can be compromised
  - Protect assets with least privilege access
  - Identity as a foundation
  - Continuous verification
- Training, Research, Resource Hub
- Full Exam & Certification available

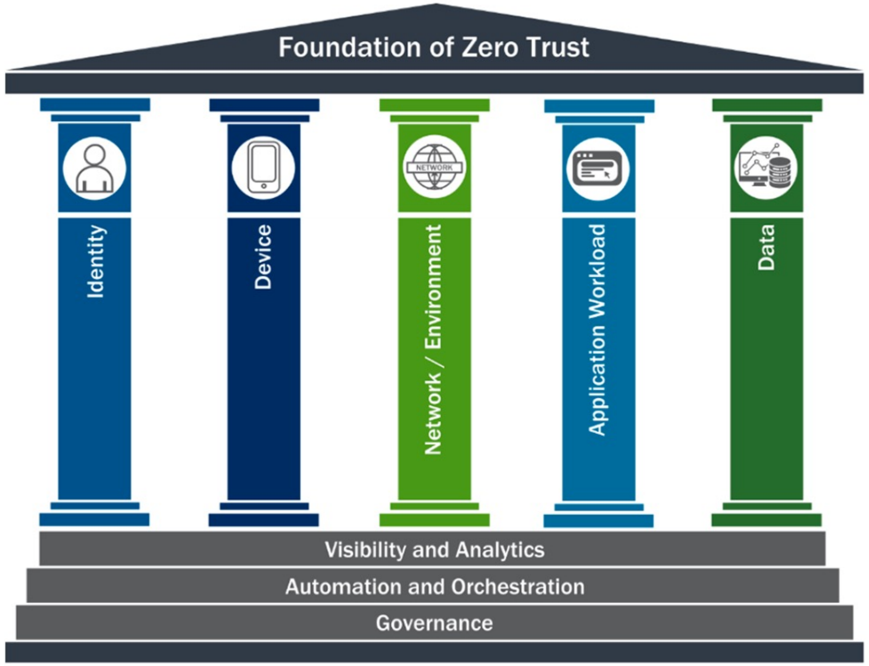
CCZT™
Certificate of
Competence in Zero Trust

Zero Trust Resource Hub

Zero Trust content provided by the industry, curated by CSA

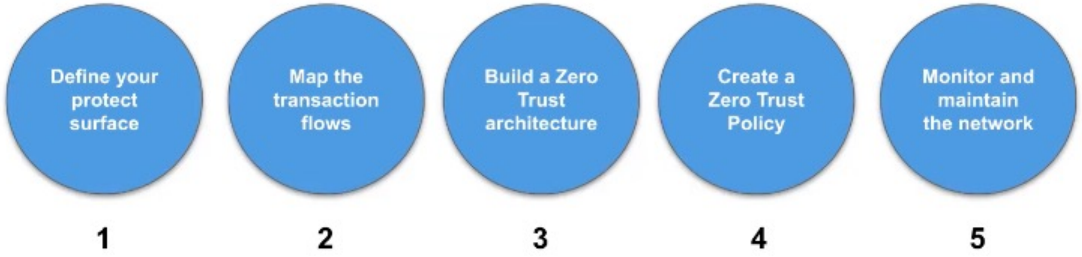
Learn More

cloud security alliance® ZERO TRUST Advancement Center

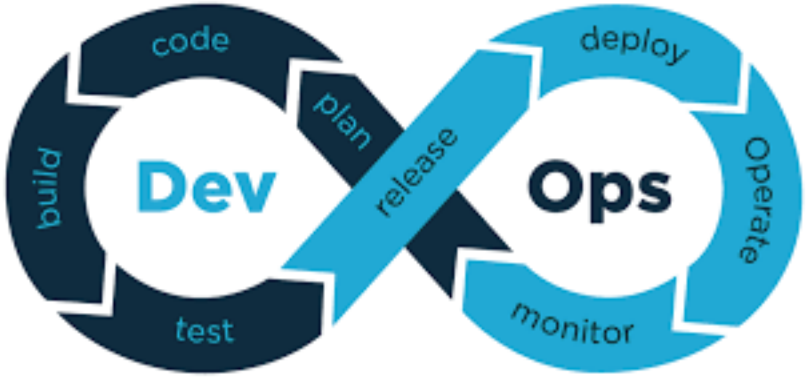


- www.cloudsecurityalliance.org/zt

- ZT Weakness is "Cloud Native"
  - 72% of containers live less than 5 minutes
  - 90% of granted permissions are not used
  - 87% of container images have high or critical vulnerabilities
  - 15% of high or critical vulnerabilities are in use
  - *Source: Sysdig Cloud Native Security & Usage Report 2023*

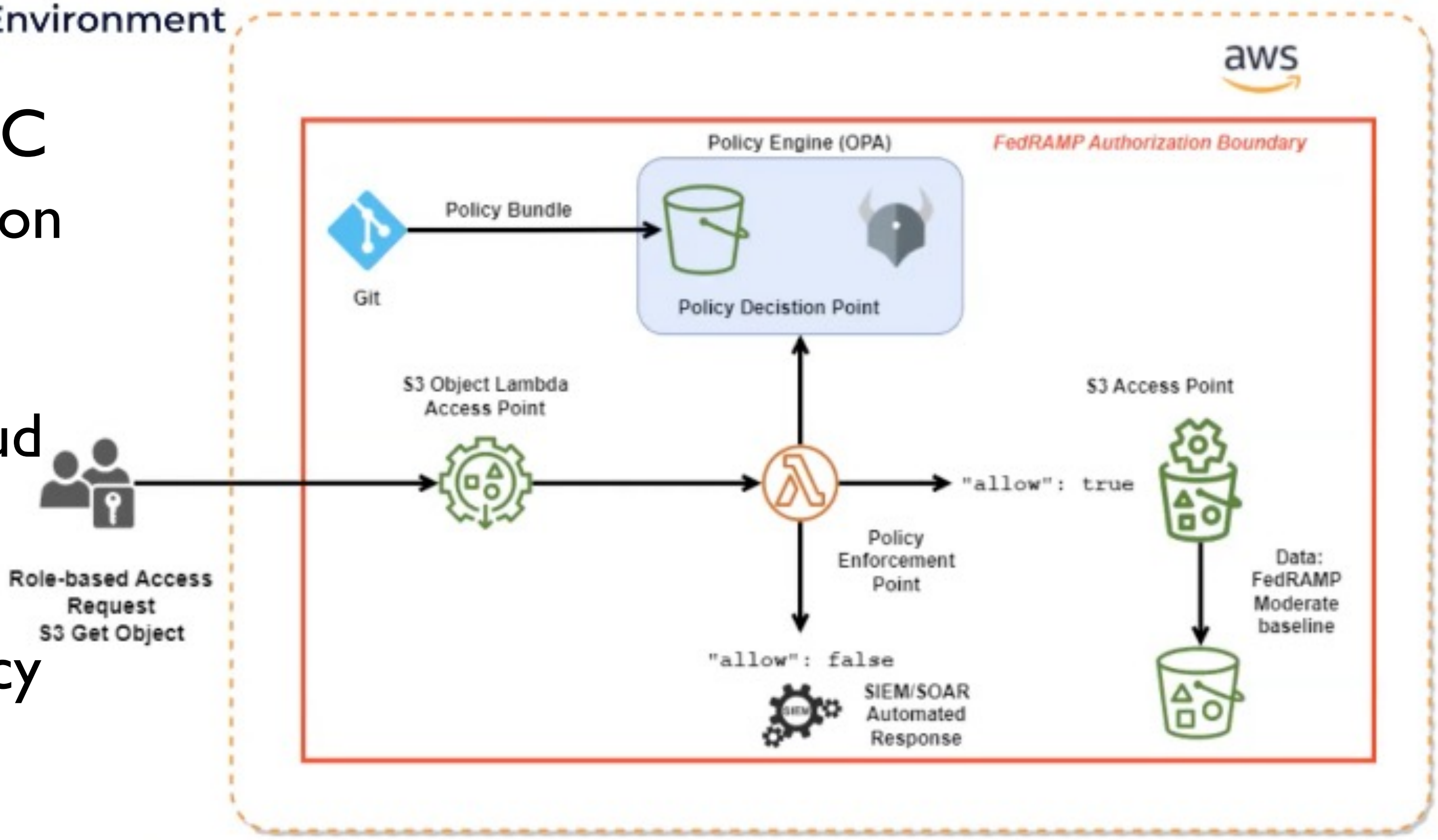

Define your protect surface
Map the transaction flows
Build a Zero Trust architecture
Create a Zero Trust Policy
Monitor and maintain the network
1
2
3
4
5

- Priority to apply ZT to DevOps & Microservices
  - Extend 5 Pillars to "Cloud Native"
  - Container and Serverless deployment
  - Focus on Automation & Tooling
  - Layer 7 Access Control
  - Infrastructure as Code
  - Policy as Code
  - Use 5 Step process in Cloud Native Process



Demo Environment

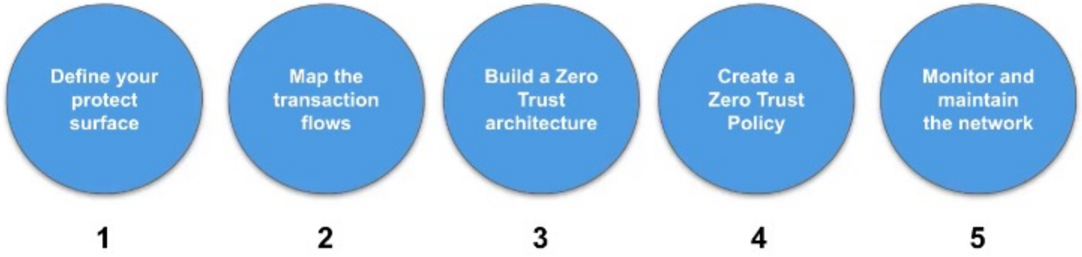
- Developed Cloud Native ZT POC
  - Protect FedRAMP Authorization Boundary containing health information
  - Developed in AWS public cloud with mock data
  - Heavy reliance on Serverless
  - Use Terraform and Open Policy Agent to orchestrate


aws
FedRAMP Authorization Boundary
Policy Engine (OPA)
Policy Bundle
Git
Policy Decision Point
S3 Object Lambda Access Point
S3 Access Point
"allow": true
Policy Enforcement Point
Data: FedRAMP Moderate baseline
"allow": false
SIEM/SOAR Automated Response
Role-based Access Request
S3 Get Object

Zero Trust should be central to AI Data Transformation & Governance

Data Transformation Steps

1. Data Collection and Integration
2. Data Cleaning and Preprocessing
3. Data Augmentation
4. Data Annotation and Labeling
5. Data Storage and Management
6. Data Privacy and Security
7. Ongoing Data Monitoring and Quality Assurance



1 Define your protect surface

2 Map the transaction flows

3 Build a Zero Trust architecture

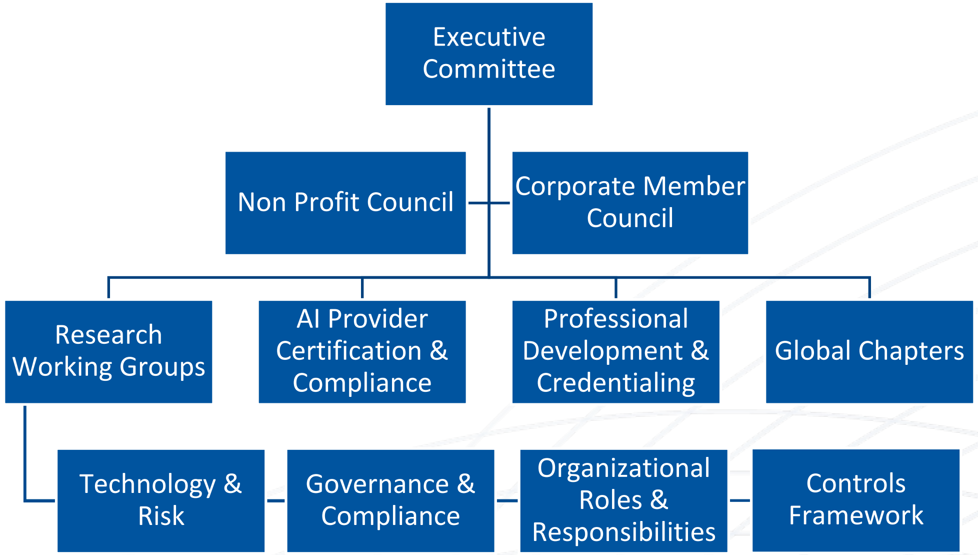
4 Create a Zero Trust Policy

5 Monitor and maintain the network

Data Governance Framework

1. Governance Structure
2. Data Policies and Standards
3. Data Quality Management
4. Data Privacy and Security
5. Data Access and Control
6. Training and Awareness
7. Technology and Tools
8. Compliance and Regulatory Requirements
9. Performance Measurement

CSA AI Safety Initiative



Executive Committee

- Frontier Model Companies
  - OpenAI
  - Google Deepmind
  - Anthropic
- Hyperscalers
  - Amazon AWS
  - Google Cloud
  - Microsoft
- CISA (USA)
- Key stakeholders from government and industry around the world

Non Profit Council

- Chaired by United Nations International Computing Centre

Corporate Member Council

- Represent 500 corporate members from AI, Cloud, Cybersecurity, Audit and all critical infrastructure industries

Additional Support

- 1,500 volunteer experts in research working groups
- Chapters from over 60 countries
- www.cloudsecurityalliance.ai

Summary

Pervasive Generative AI is the biggest technology trend to date

Cloud 3 is the merger with AI

The world is hungry for data – your data

Modern identity strategy

Robust data governance

Zero Trust is key

Thank You!

jreavis@cloudsecurityalliance.org