A secure approach to generative AI with AWS

Giuseppe Russo

Head of Security Assurance

Amazon Web Services Italy and SEE

Innovation can **transform industries**

GENERATIVE AI

More than 100,000 customers use AWS for ML

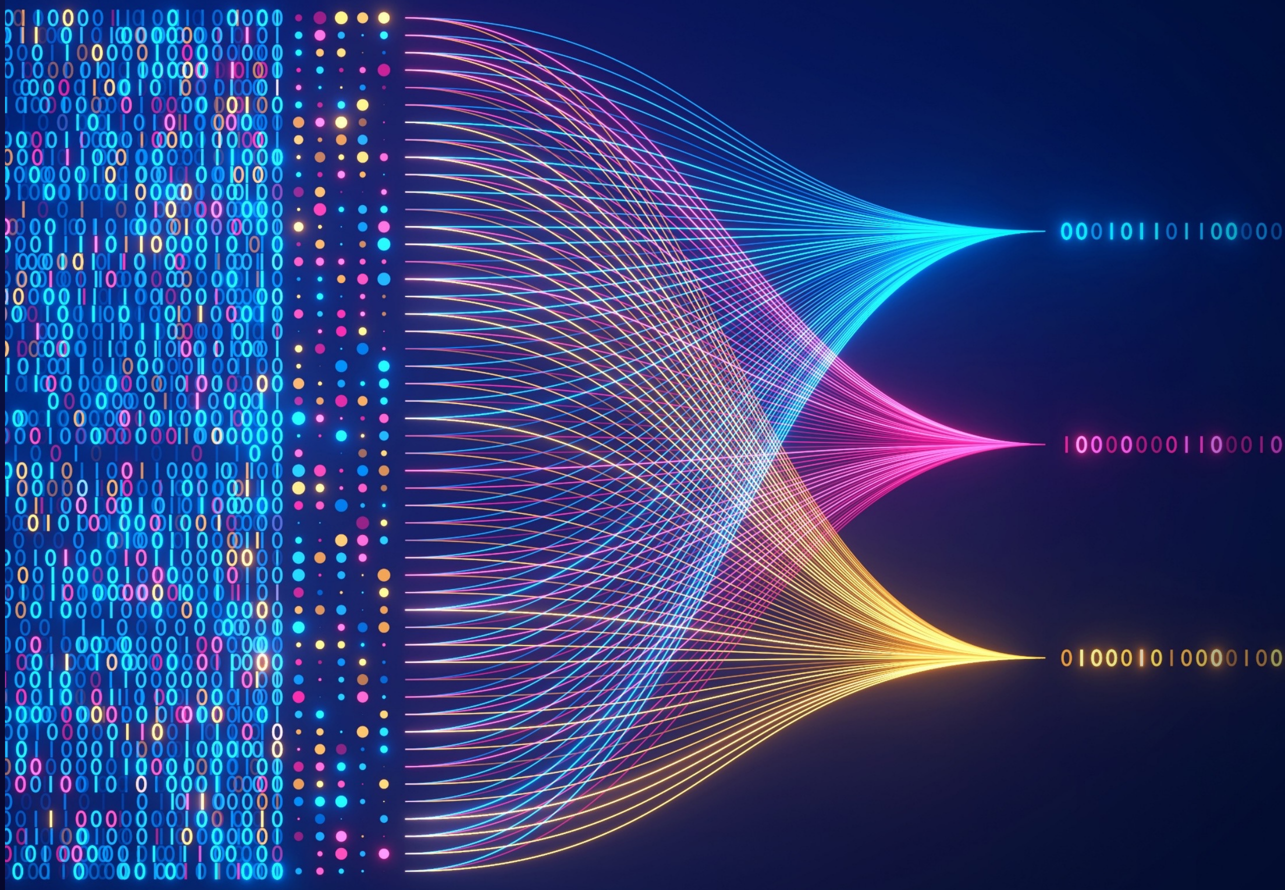
Generative AI is powered by foundation models

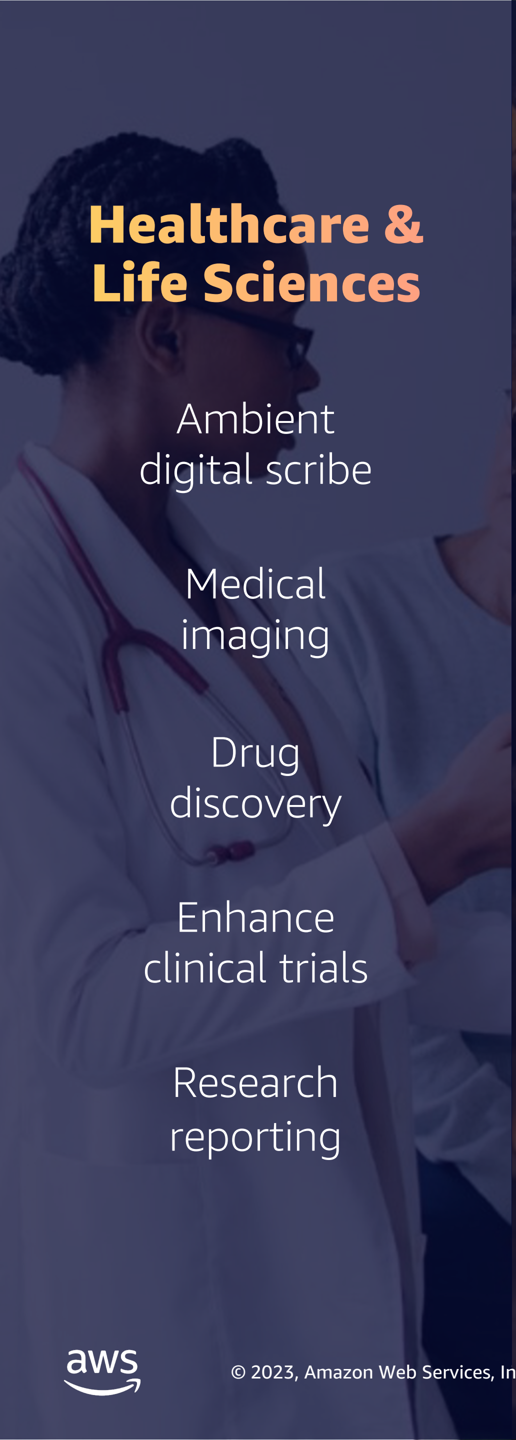
Pretrained on vast amounts of unstructured data

Contain large number of parameters that make them capable of learning complex concepts

Can be applied in a wide range of contexts

Customize FMs using your data for domain specific tasks





Healthcare & Life Sciences

Ambient digital scribe

Medical imaging

Drug discovery

Enhance clinical trials

Research reporting

Industrial & Manufacturing

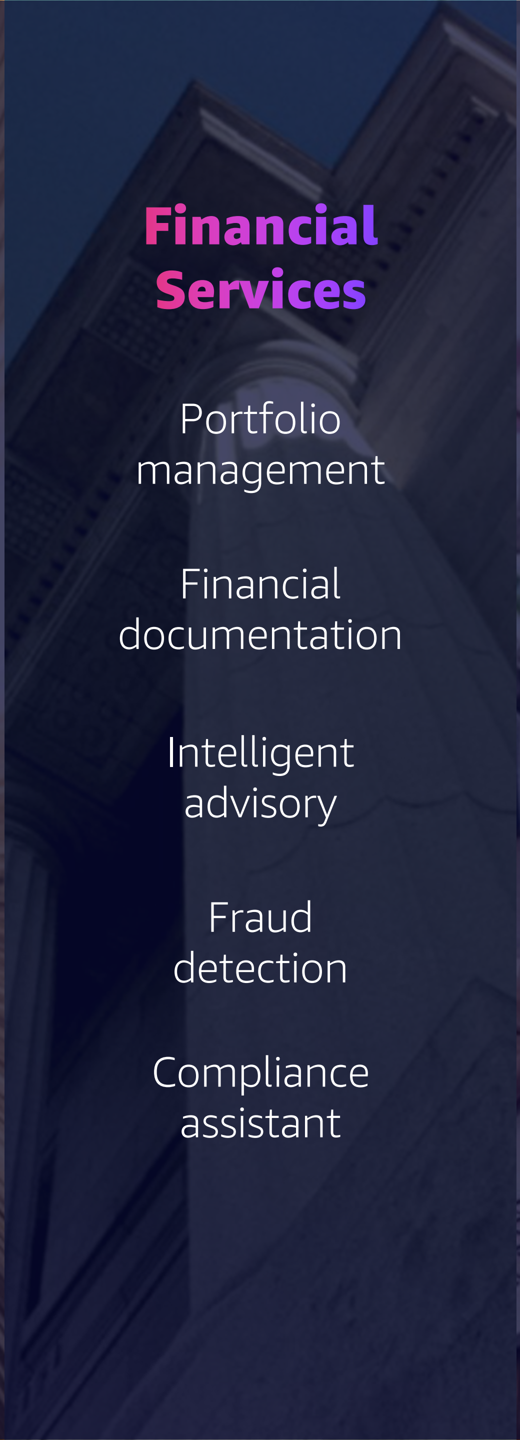
Product design

Operational efficiency

Maintenance Assistants

Supply chain optimization

Equipment diagnostics



Financial Services

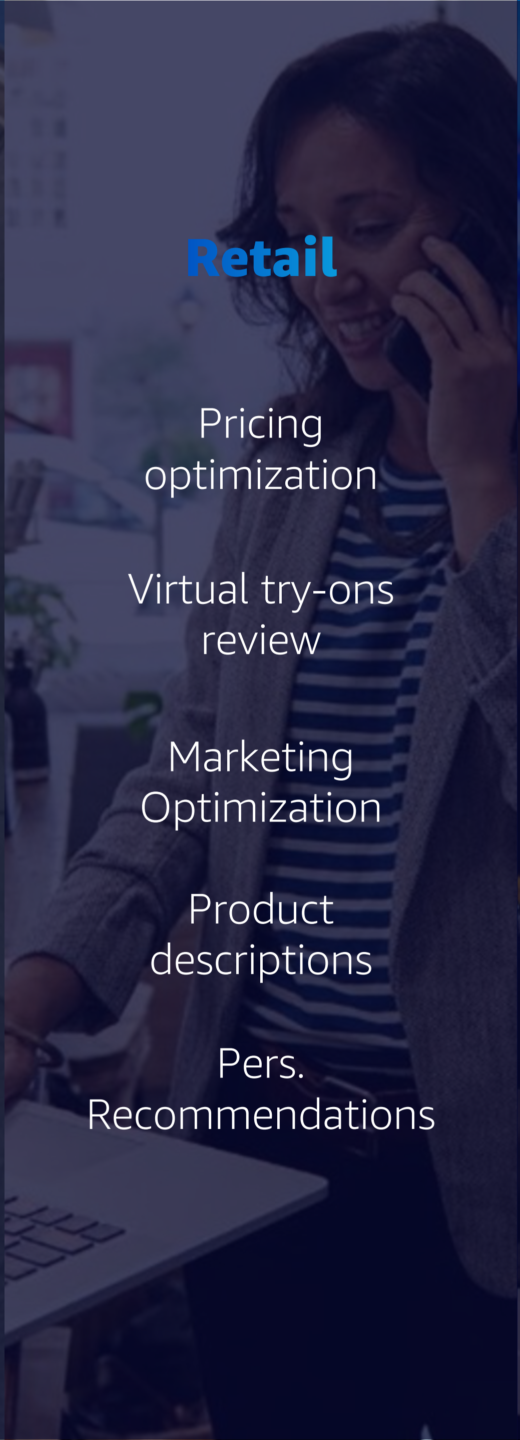
Portfolio management

Financial documentation

Intelligent advisory

Fraud detection

Compliance assistant



Retail

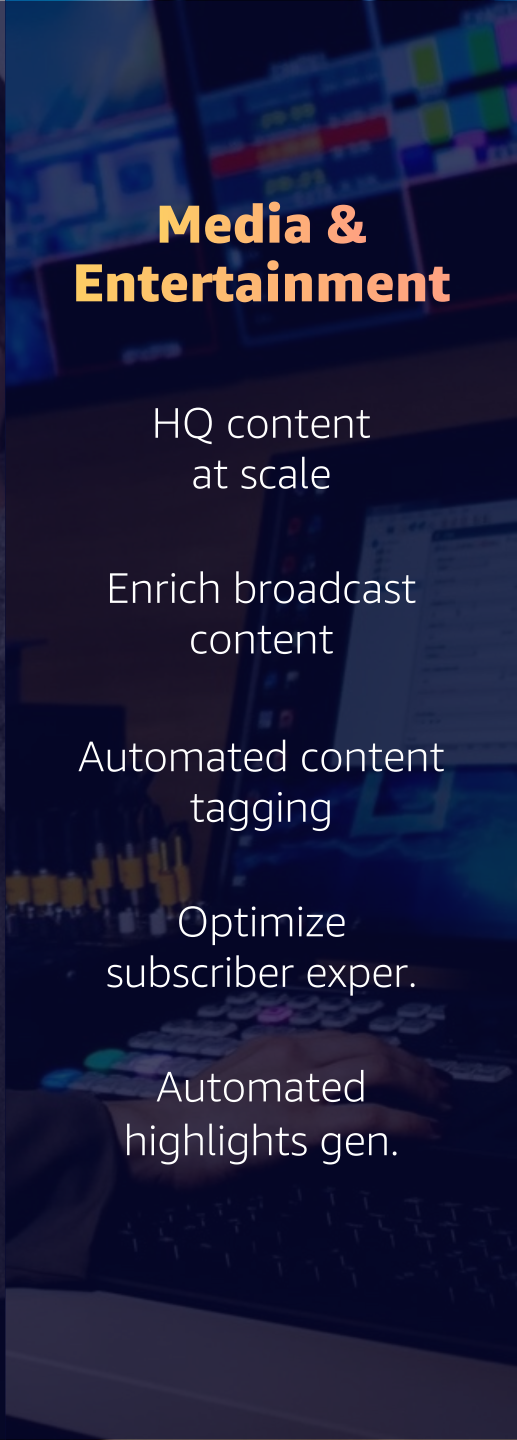
Pricing optimization

Virtual try-ons review

Marketing Optimization

Product descriptions

Pers. Recommendations



Media & Entertainment

HQ content at scale

Enrich broadcast content

Automated content tagging

Optimize subscriber exper.

Automated highlights gen.

Customers have questions…

Which model should I use?

How can I move quickly?

How can I keep my data secure & private?

Secure Deployment of Gen AI: Customer Concerns and Threats

Importance of FMs and Applications

FMs and the applications built around them represent extremely valuable investments for our customers.

They're often used with highly sensitive business data

Customer Concerns about Data Protection

CU biggest concern is how to take advantages of Gen AI, while protecting their highly sensitive data and investments.

CU data and model weights are incredibly valuable, customers require them to stay protected, secure, and private

Potential Threats to Data Security

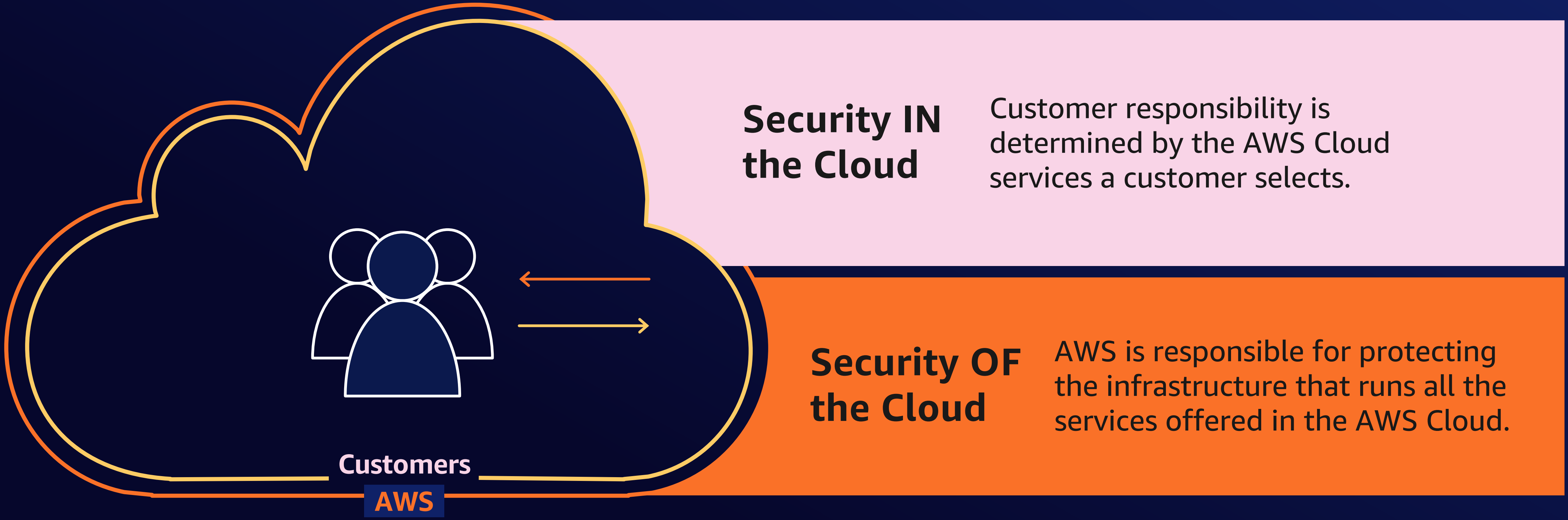
From their own administrator's accounts

From their customers

Vulnerabilities in software running in their own environments

Threats from their cloud service provider having access

AWS Shared Responsibility Model



Security IN the Cloud

Customer responsibility is determined by the AWS Cloud services a customer selects.

Security OF the Cloud

AWS is responsible for protecting the infrastructure that runs all the services offered in the AWS Cloud.

Customers

AWS

Securing Generative AI Stack

APPLICATIONS THAT LEVERAGE LLMs AND OTHER FMs

TOOLS TO BUILD WITH LLMs AND OTHER FMs

INFRASTRUCTURE FOR FM TRAINING AND INFERENCE

Protecting AI Bottom layer

APPLICATIONS THAT LEVERAGE LLMs AND OTHER FMs

TOOLS TO BUILD WITH LLMs AND OTHER FMs

INFRASTRUCTURE FOR FM TRAINING AND INFERENCE

Confidential Computing with AWS Nitro System

unparalleled computing backbone for AWS, with security and performance at its core

Engineered with a **hardware-based root of trust** using the Nitro Security Chip, allowing for the system to be continuously measured and validated

Dedicated **hardware-based virtualization** minimized attack surfaces, reducing risk of vulnerabilities

No operator or root access, including Amazon employees

VM tenancy protections. Instances are isolated from other tenants on the same EC2 host via VM, never containers.

Memory encryption enabled by default on Graviton2/3, Intel Ice Lake (TME) and AMD Milan (SME) instances

AWS Nitro System **achieved independent third-party validation**

Securing Infrastructure LLM and FM Training and Inference

Securing AI infrastructure refers to **zero access to sensitive AI data**, such as:

- **AI model**
- **AI weights**
- **data processed with those models**

by any unauthorized person, either at the infrastructure operator or at the customer

Securing AI infrastructure 3 key principles:

1. Complete **isolation** of the AI data from the infrastructure operator
2. Ability for customers to **isolate** AI data from themselves
3. **Protected** infrastructure communications

Customers have questions…

Which model should I use?

How can I move quickly?

How can I keep my data secure & private?

Amazon Bedrock

The easiest way to build and scale generative AI applications with LLMs and other FMs

Choice of industry-leading FMs from AI21 Labs, Amazon, Anthropic, Cohere, Meta, Mistral AI, and Stability AI

Customize FMs using your organization’s data

Enterprise-grade security and privacy

Secure and private

Everything you expect from an AWS service

Amazon Bedrock keeps data secure & private

None of the customer's data is used to train the underlying model

All data is encrypted in transit and at rest

Data used to customize models remains within your VPC

Support for standards, including GDPR & HIPAA

Amazon Bedrock

Recently added security capabilities

CloudWatch integration

Track usage metrics and build customized dashboards

CloudTrail integration

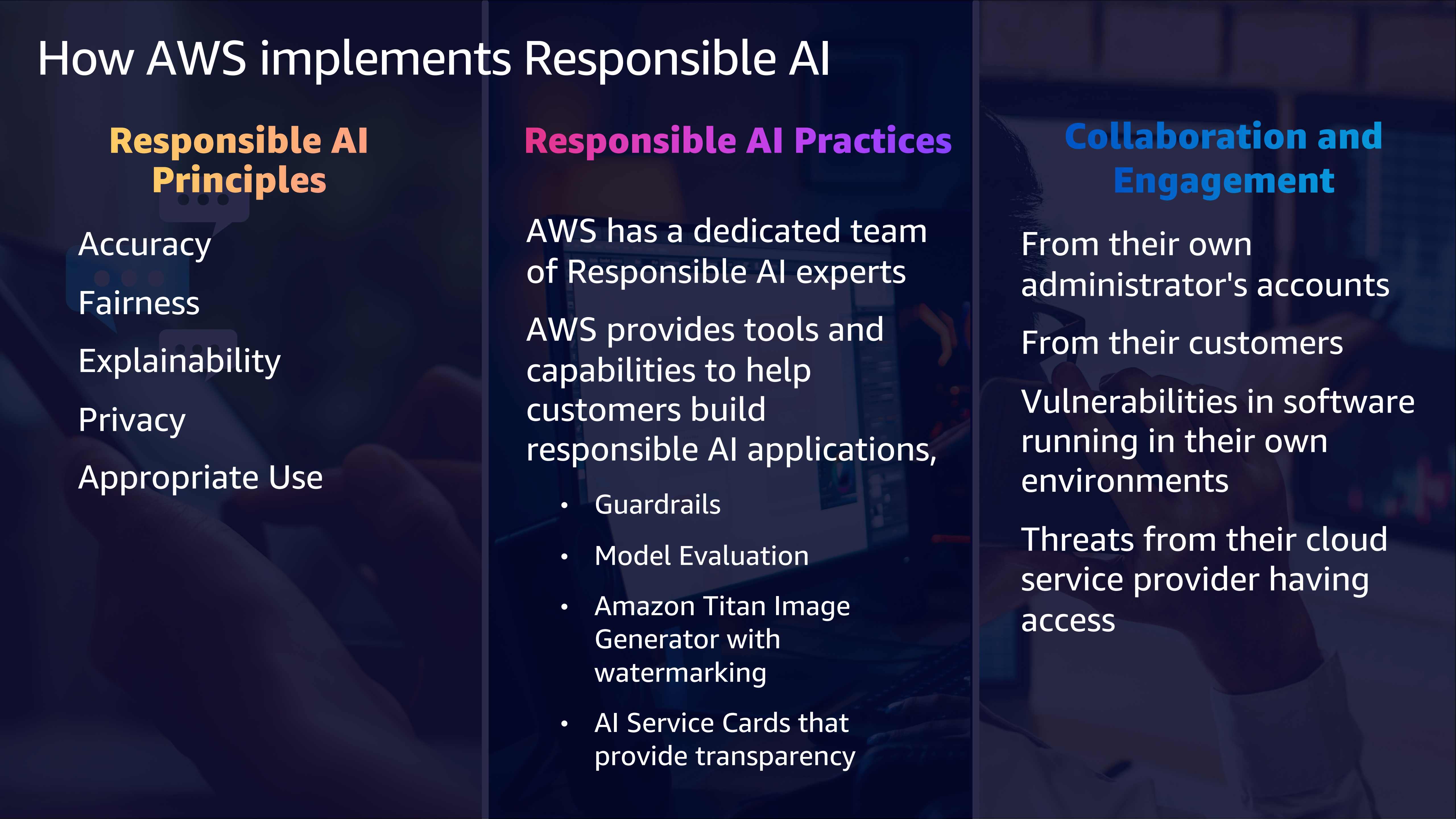
Monitor API activity and troubleshoot issues

SOC compliance

SOC 1, 2 & 3

Responsible AI

How AWS implements Responsible AI



Responsible AI Principles

Accuracy

Fairness

Explainability

Privacy

Appropriate Use

Responsible AI Practices

AWS has a dedicated team of Responsible AI experts

AWS provides tools and capabilities to help customers build responsible AI applications,

- Guardrails
- Model Evaluation
- Amazon Titan Image Generator with watermarking
- AI Service Cards that provide transparency

Collaboration and Engagement

From their own administrator's accounts

From their customers

Vulnerabilities in software running in their own environments

Threats from their cloud service provider having access