aws INNOVATE

AI/ML EDITION

aws INNOVATE
AI/ML EDITION

Accelerating machine learning innovation securely with Amazon SageMaker

Michael Stringer
Senior Security Specialist Solutions Architect
Amazon Web Services

Agenda

- Amazon SageMaker overview and new features
- Security considerations
- Demonstration of security, identity, and compliance settings
- Key points to remember

Integrated Workbench

Capabilities designed specifically for ML, data preparation, experiment management, and workflows

Managed Infrastructure

Designed for ultra low latency and high throughput, automatic scaling, and distributed training

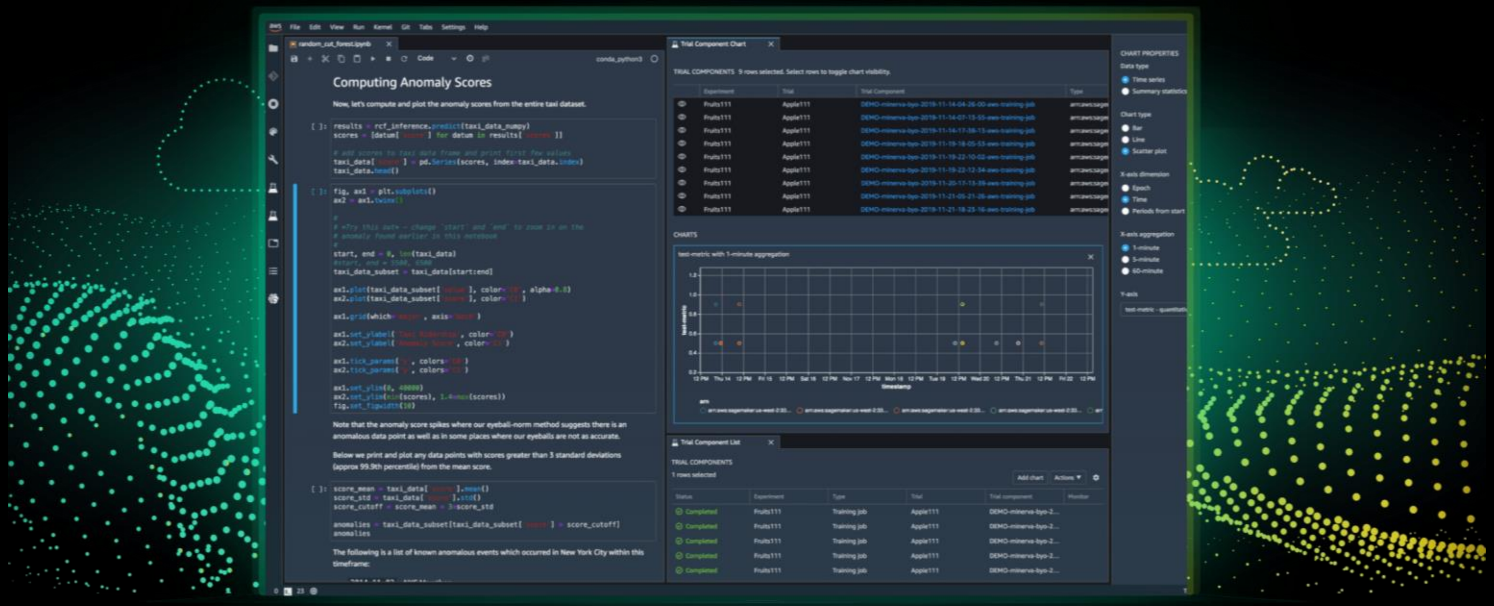
Managed Tooling

Purpose-built from the ground up to work together including auto ML, collaboration, debugger, profiler, bias analyzer, and explainability

https://aws.amazon.com/sagemaker

Amazon SageMaker

Most complete, end-to-end ML service



Computing Anomaly Scores

Now, let's compute and plot the anomaly scores from the entire taxi dataset.

```
[ ]: results = rcf_inference.predict(taxi_data_numpy)
     scores = [datum['score'] for datum in results['scores']]

     # add scores to taxi data frame and print first few values
     taxi_data['score'] = pd.Series(scores, index=taxi_data.index)
     taxi_data.head()
```

```
[ ]: fig, ax1 = plt.subplots()
     ax2 = ax1.twinx()

     #
     # *Try this out* - change `start` and `end` to zoom in on the
     # anomaly found earlier in this notebook
     #
     start, end = 0, len(taxi_data)
     #start, end = 5500, 6500
     taxi_data_subset = taxi_data[start:end]

     ax1.plot(taxi_data_subset['value'], color='C0', alpha=0.8)
     ax2.plot(taxi_data_subset['score'], color='C1')

     ax1.grid(which='major', axis='both')

     ax1.set_ylabel('Taxi Ridership', color='C0')
     ax2.set_ylabel('Anomaly Score', color='C1')

     ax1.tick_params('y', colors='C0')
     ax2.tick_params('y', colors='C1')

     ax1.set_ylim(0, 40000)
     ax2.set_ylim(min(scores), 1.4*max(scores))
     fig.set_figwidth(10)
```

Note that the anomaly score spikes where our eyeball-norm method suggests there is an anomalous data point as well as in some places where our eyeballs are not as accurate.

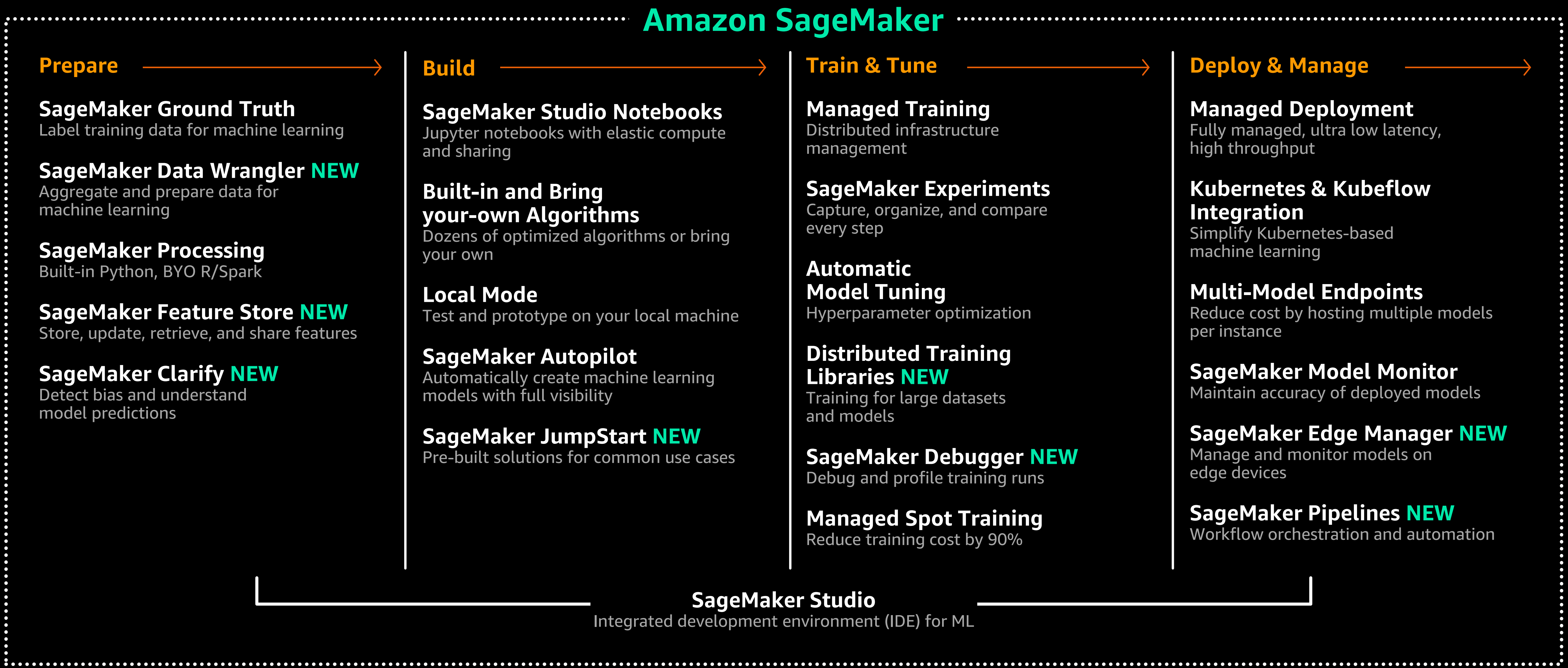
Below we print and plot any data points with scores greater than 3 standard deviations (approx 99.9th percentile) from the mean score.

```
[ ]: score_mean = taxi_data['score'].mean()
     score_std = taxi_data['score'].std()
     score_cutoff = score_mean + 3*score_std

     anomalies = taxi_data_subset[taxi_data_subset['score'] > score_cutoff]
     anomalies
```

The following is a list of known anomalous events which occurred in New York City within this timeframe:

Amazon SageMaker overview



Amazon SageMaker

Prepare

SageMaker Ground Truth
Label training data for machine learning

SageMaker Data Wrangler NEW
Aggregate and prepare data for machine learning

SageMaker Processing
Built-in Python, BYO R/Spark

SageMaker Feature Store NEW
Store, update, retrieve, and share features

SageMaker Clarify NEW
Detect bias and understand model predictions

Build

SageMaker Studio Notebooks
Jupyter notebooks with elastic compute and sharing

Built-in and Bring your-own Algorithms
Dozens of optimized algorithms or bring your own

Local Mode
Test and prototype on your local machine

SageMaker Autopilot
Automatically create machine learning models with full visibility

SageMaker JumpStart NEW
Pre-built solutions for common use cases

Train & Tune

Managed Training
Distributed infrastructure management

SageMaker Experiments
Capture, organize, and compare every step

Automatic Model Tuning
Hyperparameter optimization

Distributed Training Libraries NEW
Training for large datasets and models

SageMaker Debugger NEW
Debug and profile training runs

Managed Spot Training
Reduce training cost by 90%

Deploy & Manage

Managed Deployment
Fully managed, ultra low latency, high throughput

Kubernetes & Kubeflow Integration
Simplify Kubernetes-based machine learning

Multi-Model Endpoints
Reduce cost by hosting multiple models per instance

SageMaker Model Monitor
Maintain accuracy of deployed models

SageMaker Edge Manager NEW
Manage and monitor models on edge devices

SageMaker Pipelines NEW
Workflow orchestration and automation

SageMaker Studio
Integrated development environment (IDE) for ML

Amazon SageMaker Studio

Fully Integrated Development Environment (IDE) for machine learning

Collaboration at scale

Share notebooks without tracking code dependencies

Easy experiment management

Organize, track, and compare thousands of experiments

Automatic model generation

Get accurate models with full visibility and control without writing code

Higher quality ML models

Automatically debug errors, monitor models, and maintain high quality

Increased productivity

Code, build, train, deploy, and monitor in a unified visual interface

Amazon SageMaker Notebooks

Fast-start sharable notebooks

Easy access with Single Sign-On (SSO)

Access your notebooks in seconds

Fully managed and secure

Administrators manage access and permissions

Fast setup

Start your notebooks without spinning up compute resources

Easy collaboration

Share notebooks with a single click

Flexible

Dial up or down compute resources (coming soon)

Amazon SageMaker Pipelines

Managed machine learning CI/CD service

Centrally **manage** each step of the workflow

Share and re-run workflows

Choose from **built-in** templates

Compare workflows visually

Amazon SageMaker Feature Store: Securely store, discover, and share features for machine learning

Online and offline

Millisecond latency

Consistent features

Visually search

Sharing and collaboration

Balancing ML agility with IT governance

ML Builders

Focus on unique business value

Self-service access

Experiment fast

Respond quickly to change

Cloud IT and DevOps

Security

Compliance

Operations

Spend management

Amazon SageMaker security features help you go from idea to production faster

Infrastructure and network isolation
Control data traffic across Amazon SageMaker components over a private network, and ensure appropriate ingress/egress with single-tenancy

Authentication and authorization
Define, enforce, and audit who can be authenticated and authorized to use Amazon SageMaker resources

Data protection
Ensure automatic data encryption at rest and in transit with flexibility to bring your own keys

Monitoring and auditability
Track, trace, and audit all application programming interface (API) calls, events, data access, or interactions down to the user and IP level to ensure quick remediation

Compliance certifications
Inherit the most comprehensive compliance controls, and easily abide by your industry's legislation

SECURE ML WITH AMAZON SAGEMAKER

Infrastructure & Network Isolation

Control data traffic across SageMaker components over a private network, and ensure appropriate ingress/egress with single-tenancy

Amazon Virtual Private Cloud (VPC) Support

Deploy Amazon SageMaker Instances with an Elastic Network Interface (ENI) connected to a private subnet within your Virtual Private Cloud (VPC), and create security groups to only allow the inbound and outbound traffic that is required for your use case(s).

Network Isolation with optional AWS PrivateLink Support

Control your network traffic via your VPC, or completely disable internet egress. Optionally leverage AWS PrivateLink for advanced private connectivity.

Multi-tenant Isolation

Ensure ML environment single-tenancy as Amazon SageMaker instances are deployed on single-tenancy Amazon Elastic Compute Cloud (EC2) instances in the service platform.

Root Access Management

Manage user root access in a programmatic fashion, and decide when to give your data scientists the flexibility they need to leverage external libraries, or when to solely allow access to available libraries on the AWS CodeArtifact.

SECURE ML WITH AMAZON SAGEMAKER

Authentication & Authorization

Define, enforce, and audit who can be authenticated (signed in) and authorized (have permissions) to use Amazon SageMaker resources

IAM Role-based Access Controls

Map users/groups/roles from on-prem AD/LDAP to AWS Identity and Access Management (IAM) roles and control which SageMaker features those users and roles have access to.

Multi-factor Authentication

SageMaker notebooks are accessed from the AWS console which if set up provides Multi-factor Authentication (MFA) support.

Tag-based Access Control

Add tags to your Amazon SageMaker resources, and provide the tag information in the condition element of a policy using condition keys to control access.

Detective Controls

Implement intelligent threat detection and continuous monitoring to protect from unauthorized actions or access to your ML environment and data.

Preventive Controls

Prevent specific actions that are taken by principals in your environment to ever succeed unless certain conditions are met.

SECURE ML WITH AMAZON SAGEMAKER

Data Protection

Ensure automatic data encryption at rest and in transit with flexibility to bring your own keys

End-to-end Encryption at Rest

Encrypt model artifacts and data in Amazon Simple Storage Service (S3), root volume, and Amazon Elastic Block Store (EBS) volume across notebooks, training jobs, and inference endpoints.

End-to-end Encryption in Transit

Encrypt all inter-network communications, and ensure that Amazon SageMaker API calls are made over a secure (SSL) connection with the necessary IAM permissions.

Inter-container/network Encryption

Enhance your encryption by requiring all communications between training nodes of a distributed training cluster to be encrypted.

Encryption with Customer Managed Keys

Use your own customer managed keys to encrypt and decrypt your data across Amazon S3 input/output buckets, as well as root and Amazon EBS volumes.

SECURE ML WITH AMAZON SAGEMAKER

Monitoring & Auditability

Track, trace, and audit all API calls, events, data access, or interactions down to the user and IP level to ensure quick remediation

Activity Logging and Event Collection

Monitor, store, and access your log files from Amazon EC2 instances, AWS CloudWatch, and other sources, and build custom dashboards or set alarms to take remedial actions based on specified thresholds on your customer metrics.

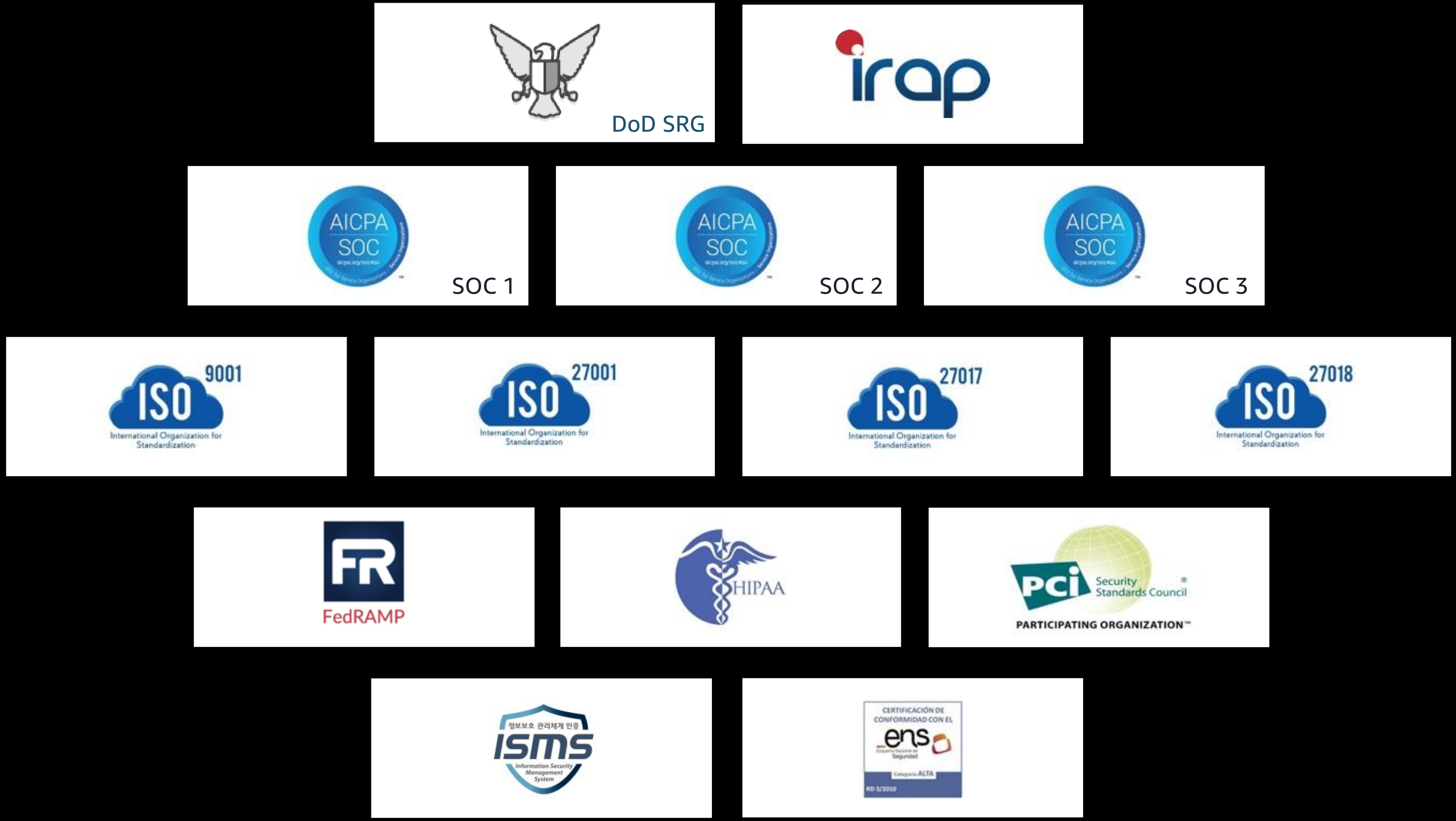
AWS CloudTrail & Amazon CloudWatch Audits

Track user activity and API usage to enable governance, compliance, and operational/risk auditing of your ML environment.

SECURE ML WITH AMAZON SAGEMAKER

Compliance Certifications

Inherit the most comprehensive compliance controls, and easily abide by your industry’s legislation


DoD SRG
irap
AICPA SOC
SOC 1
AICPA SOC
SOC 2
AICPA SOC
SOC 3
ISO 9001
International Organization for Standardization
ISO 27001
International Organization for Standardization
ISO 27017
International Organization for Standardization
ISO 27018
International Organization for Standardization
FR
FedRAMP
HIPAA
PCI Security Standards Council
PARTICIPATING ORGANIZATION™
ISMS
Information Security Management System
CERTIFICACIÓN DE CONFORMIDAD CON EL
ens
Categoría ALTA

Amazon SageMaker security controls

Infrastructure and network isolation

VPCs, Security Groups, PrivateLink connections

Authentication and authorization

AWS IAM and AWS Single Sign On (SSO); detective and preventative controls with AWS Config, Amazon GuardDuty, and AWS Security Hub

Data protection

TLS encryption in transit; AWS Key Management Service (KMS) encryption at rest

Monitoring and auditability

AWS CloudTrail and Amazon CloudWatch

Compliance certifications

AWS compliance with more than 60 regulatory frameworks

aws INNOVATE
AI/ML EDITION

Demo

Key security points to remember

- Security controls are ‘baked in’ to Amazon SageMaker
- Implementing controls is easy
- Amazon SageMaker is fully integrated with services to continuously monitor and audit usage
- AWS provide services for detective and preventative controls
- Security, identity, and compliance can be done at the ‘speed of Cloud’!

Visit the AI and Machine Learning Resource Hub for more resources

Dive deeper with these resources, get inspired and learn how you can use machine learning to accelerate business outcomes.

- The machine learning journey e-book
- Machine learning enterprise guide
- 7 leading machine learning use cases e-book
- A strategic playbook for data, analytics, and machine learning
- Accelerating ML innovation through security e-book
- … and more!

Visit resource hub »

https://tinyurl.com/aiml-aws

AWS Machine Learning (ML) Training and Certification

Learn like an Amazonian, based on the curriculum we’ve used to train our own developers and data scientists

AWS is how you build machine learning skills

Courses built on the curriculum leveraged by Amazon’s own teams. Learn from the experts at AWS.

Flexibility to learn your way

Learn online with 65+ on-demand digital courses or live with virtual instructor-led training, plus hands-on labs and opportunities for practical application.

Validate your expertise

Demonstrate expertise in building, training, tuning, and deploying machine learning models with an industry-recognized credential.

aws.training/machinelearning

Thank You for Attending AWS Innovate

We hope you found it interesting! A kind reminder to **complete the survey.** Let us know what you thought of today's event and how we can improve the event experience for you in the future.

- aws-apac-marketing@amazon.com
- twitter.com/AWSCloud
- facebook.com/AmazonWebServices
- youtube.com/user/AmazonWebServices
- slideshare.net/AmazonWebServices
- twitch.tv/aws

aws

aws INNOVATE
AI/ML EDITION

Thank you!