Accelerating machine learning innovation through security

Security features from Amazon SageMaker and the AWS Cloud can help you go from idea to production faster.

INTRODUCTION

How security helps deliver machine learning results

To build successful machine learning models, you often need datasets unique to your business. These datasets are extremely valuable assets and need to be secured throughout every step of the machine learning process—including data preparation, training, validation, and inference.

In a typical machine learning project, it can take months to build a secure workflow before you can begin any work on your models. Maintaining executive buy-in means delivering fast results— so accelerating projects by weaving security into every step of the process will help ensure organization-wide commitment to your project and your larger machine learning initiatives.

Amazon SageMaker is a fully managed service that provides every developer and data scientist with the ability to build, train, and deploy machine learning models quickly and securely. In this eBook, we provide an overview of the Amazon SageMaker security features that can help your organization meet the strict security requirements of machine learning workloads—ultimately helping you go from idea to production faster, more securely, and with a higher rate of success.

Executive summary

As a managed AWS service, Amazon SageMaker automatically inherits the AWS global infrastructure and its network security features. AWS is purpose built for the cloud, with data centers and a network architected to help protect your information, identities, applications, and devices. The AWS network and infrastructures are monitored 24/7 to ensure confidentiality, integrity, and availability of your data. In addition, Amazon SageMaker offers a comprehensive set of capabilities, so you can run your machine learning workloads with the most flexible and secure machine learning environment available today.

Customers have told us that the following are the key security criteria they consider when evaluating machine learning solutions. Together, AWS Cloud and Amazon SageMaker security features allow you to meet these criteria readily—so you can put machine learning to work securely in production applications.

Infrastructure and network security

Control data traffic across Amazon SageMaker components over a private network. Ensure appropriate ingress/egress with single-tenancy, so your data and resources are secure.

Data protection

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

Compliance certifications

Inherit the most comprehensive compliance controls and easily meet your industry's regulatory requirements.

Authentication and authorization

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

Monitoring and auditability

Track, trace, and audit all API calls, events, data access, and interactions down to the user and IP levels.

Infrastructure and network security

Machine learning security starts with the core infrastructure, including underlying compute, storage, and networking. When assessing infrastructure and network security of machine learning solutions, look for these critical qualifications: 1) the ability to isolate the network and keep data traffic across the various components of the workflow within secure private network connections; 2) the ability to control access, and, more specifically, to block inflow (ingress) and outflow (egress) of data and code from and to the internet; and 3) a tenancy model that provides isolation between user environments.

Amazon SageMaker uses Amazon Virtual Private Cloud (VPC), a service that provides logically isolated sections of the AWS Cloud to launch its resources in a virtual network of its own. All data traffic between various Amazon SageMaker components flows within this network, controlled tightly by security group permissions. You also have the option to deploy Amazon SageMaker within your own VPC to provide secure access to your private resources. In addition, Amazon SageMaker enables network isolation from the internet by allowing you to disable outbound data traffic to the internet through its network. This option helps prevent users from engaging in risky behaviors, such as installing unauthorized software.

You can also control Amazon SageMaker's network traffic using AWS PrivateLink, a service that provides private connectivity between VPCs, AWS services, and on-premises applications. Further, Amazon SageMaker instances are deployed on single-tenancy Amazon EC2 instances to ensure that your machine learning environments are isolated from other customers. Lastly, Amazon SageMaker allows you to restrict root access to users in a programmatic fashion, so you can decide when to give your data scientists the flexibility they need to leverage external libraries.

Learn more about infrastructure security in Amazon SageMaker ›

3M innovates while maintaining focus

Through research and development, 3M introduces more than 1,200 new products every year. The company is using Amazon SageMaker to improve the effectiveness of its quality control processes and move beyond time-consuming manual inspection. 3M built machine learning models to improve materials research, analysis, and defect detection.

> **"Our new machine learning-based processes are far more efficient than prior approaches. (Amazon) SageMaker provided cost-effective access to powerful infrastructure on demand, along with comprehensive security features. This allowed us to focus on research rather than the mechanics of securely scaling our compute capabilities."**
>
> David Frazee, Vice President, 3M Corporate Research Systems Lab

Authentication and authorization

One of the fundamental capabilities you need to secure your machine learning environment is a strong mechanism to define, enforce, and audit who can sign in (called authentication) and what resources and functions they are authorized to access (called authorization).

Amazon SageMaker is governed by AWS Identity and Access Management (IAM), a service that enables you to manage access to AWS services and resources securely. With AWS IAM, you can implement fine-grained access controls. AWS IAM allows you to specify who can perform what actions to which resources and under what circumstances at the level of specific features, users, groups, and roles. You can readily bring existing user identities from AWS Directory Service, your enterprise user directory such as Active Directory (AD), Lightweight Directory Access Protocol (LDAP), or a web identity provider.

1

Multi-factor authentication (MFA), which prompts users for their user name and password (the first factor), and an authentication code from their AWS MFA device (the second factor)

2

Tag-based access control to categorize resources by purpose, owner, environment, and other criteria, making it easier to manage, search, and filter resources

3

Detective controls that identify potential security threats or incidents using user behavior in Amazon SageMaker

4

Preventive controls that can stop a potentially harmful action before it takes place

Learn more about AWS IAM for Amazon SageMaker ›

Data protection

Another important security requirement for machine learning solutions is protecting data through automatic encryption at rest, in transit, and during training across distributed clusters. Machine learning solutions should also provide the flexibility to bring your own encryption keys.

Amazon SageMaker comes with built-in encryption capabilities to ensure that training datasets, input data for inference, and other machine learning model and system artifacts are encrypted in transit and at rest. Amazon SageMaker also gives you flexible data encryption options through Amazon SageMaker managed keys, AWS managed keys, and customer managed

Learn more about data protection in Amazon SageMaker ›

The NFL tackles player safety

Together, the NFL and AWS are leveraging machine learning to build the "Digital Athlete," a platform to improve injury prevention and treatment—and, ultimately, predict injury. The program will use anonymized and aggregated player data to create a composite that will simulate infinite scenarios of the game environment. The NFL and AWS hope the program will eventually have implications beyond football—for example, it could become a useful tool in the healthcare industry.

> **"Since the data used in the modeling is highly sensitive, we needed an ML solution like Amazon SageMaker with security and compliance features built-in to protect the data throughout the ML process."**
>
> Jennifer Langton, SVP Health and Safety, NFL

Monitoring and auditability

Auditability is about tracking, tracing, and monitoring API calls, events, data access, and interactions down to the user and IP levels to ensure quick remediation (if necessary). It's critical to be able to capture audit trails at the granular level of users, files, and objects.

Amazon SageMaker is integrated with Amazon CloudWatch Logs and AWS CloudTrail for logging events and API calls. You can also set alarms that watch for certain thresholds and send notifications or take actions when those thresholds are met. And you can identify which users and accounts called AWS, the source IP address from which the calls were made, and when the calls occurred. Since Amazon SageMaker uses data from Amazon Simple Storage Service (Amazon S3), all data access activities are automatically logged for monitoring.

Learn more about logging and monitoring in Amazon SageMaker ›

"We chose (Amazon) CloudWatch because it has simplified the process to aggregate various types of log streams from various AWS services, and reduced the overhead costs and complexity generally associated with logs management. (Amazon) CloudWatch also supports the ability to export certain logs via log streams—which enabled us to set up an external automated log analysis pipeline—so it was a clear winner."

Moe Abbas, Cloud Team Lead, Canva

Regulatory compliance

In many cases, machine learning solutions need to comply with regulatory standards and pass compliance certifications that vary significantly across countries and industries.

AWS supports more security standards and compliance certifications than any other cloud vendor. As an AWS service, Amazon SageMaker complies with a wide range of compliance programs, including PCI, HIPAA, SOC 1/2/3, FedRAMP, and ISO 9001/27001/27017/27018. In addition, to aid your compliance efforts, AWS regularly achieves third-party validation for thousands of global compliance requirements across finance, retail, healthcare, government, and more. For the latest SageMaker certifications, see the **AWS Compliance Program website**.

THOMSON REUTERS

Thomson Reuters innovates faster with security

Thomson Reuters is a leading source of intelligent and trusted information for businesses and professionals. Using Amazon SageMaker, Thomson Reuters accelerated the development of machine learning models for a number of innovative solutions—including text classification and natural language question answering—with cost savings and flexibility. To complement Amazon SageMaker's security features, the company built a custom solution it calls Secure Content Workspaces (SCW).

"Amazon SageMaker saved our team countless hours of coding that would have been necessary on self-managed ML infrastructure. Together, Amazon SageMaker and SCW make it possible for research and data scientists to work in the cloud in compliance with our standards without being cloud experts."

John Duprey, Senior Director of Engineering, Center for AI and Cognitive Computing, Thomson Reuters

Try Amazon SageMaker for two months, free

Amazon SageMaker can help your organization secure your machine learning environment quickly—so you can focus on scaling and innovating faster. As part of the AWS Free Tier, you can get started with Amazon SageMaker for free.

Start your free trial ›

aws