Manual de Estudio: AWS Certified AI Practitioner

Dominio 5.1: Seguridad y Protección de Sistemas de IA

Material Técnico Detallado

1. Modelo de Responsabilidad Compartida

La seguridad en AWS se rige por la división de tareas entre el proveedor y el cliente.

- **Responsabilidad de AWS (Seguridad DE LA nube):** Protección de la infraestructura global (Regiones, AZs), hardware físico, centros de datos, sistemas operativos host y capas de virtualización.
- **Responsabilidad del Cliente (Seguridad EN LA nube):** Configuración de servicios, gestión de parches en instancias (EC2), seguridad de las aplicaciones, cifrado de datos y gestión de identidades.

2. Gestión de Identidades y Accesos (IAM)

AWS Identity and Access Management (IAM) es el servicio gratuito para administrar permisos y acceso a los recursos.

2.1. Identidades y Entidades

- **Usuario Raíz (Root User):** Acceso total. **Práctica recomendada:** No usar para tareas diarias, activar MFA inmediatamente y eliminar sus claves de acceso.
- **Usuarios de IAM:** Personas u aplicaciones con credenciales de larga duración.
- **Grupos de IAM:** Colección de usuarios para simplificar la asignación de permisos por función laboral (ej. científicos de datos).
- **Roles de IAM:** Identidades temporales que pueden ser asumidas por usuarios o servicios de AWS. Proporcionan credenciales temporales que caducan automáticamente.

2.2. Políticas de IAM

Documentos en formato **JSON** que definen permisos.

- **Principio de Mínimo Privilegio:** Conceder solo los permisos necesarios para completar la tarea.
- **Denegación Explícita:** Una instrucción *Deny* siempre anula cualquier instrucción *Allow*.

3. Seguridad de Datos y Cifrado

3.1. Tipos de Cifrado

1. **Cifrado en reposo:** Datos almacenados. SageMaker cifra por defecto los volúmenes de entrenamiento.
2. **Cifrado en tránsito:** Datos moviéndose por la red. Todas las APIs de AWS soportan TLS (HTTPS).

3.2. Servicios de Protección

- **AWS Key Management Service (KMS):** Permite crear y controlar claves de cifrado. Las claves pueden ser administradas por AWS o por el cliente.
- **Amazon Macie:** Utiliza ML para identificar y alertar sobre la presencia de PII (Información de Identificación Personal) en Amazon S3.

4. Vulnerabilidades Específicas de la IA

Los sistemas de IA presentan vectores de ataque únicos que deben mitigarse proactivamente:

Amenazas a la IA

- **Envenenamiento de Datos (Data Poisoning):** Inyectar datos maliciosos en el conjunto de entrenamiento para alterar el comportamiento del modelo.
- **Entradas Contradictorias (Adversarial Inputs):** Manipulaciones sutiles en la entrada (ej. píxeles en una foto) para provocar clasificaciones erróneas.
- **Inversión de Modelo:** El atacante estudia las respuestas del modelo para deducir datos privados del entrenamiento.
- **Inyección de Peticiones (Prompt Injection):** Instrucciones maliciosas enviadas a un LLM para saltar sus reglas de seguridad.

5. Gobernanza y Trazabilidad en Amazon SageMaker

Para cumplir con requisitos regulatorios, AWS ofrece herramientas de auditoría y control.

- **AWS CloudTrail:** Registra todas las llamadas a las APIs de AWS (quién, cuándo y desde dónde).
- **SageMaker Model Cards:** Documentación inmutable del modelo (usos previstos, riesgos, métricas).

- **SageMaker Model Registry:** Catálogo con control de versiones para gestionar el ciclo de vida del modelo (aprobado, rechazado, pendiente).
- **SageMaker Model Monitor:** Supervisa en tiempo real la calidad del modelo y los datos en producción para detectar desviaciones (*drifts*).
- **Seguimiento de Trazabilidad (Lineage Tracking):** Representación gráfica de los elementos del flujo de trabajo de ML (qué dataset creó qué modelo).

6. Seguridad de Red

- **Amazon VPC:** Aísla lógicamente la infraestructura de red del cliente.
- **AWS PrivateLink:** Permite que SageMaker Studio se comunique con otros servicios (S3, CloudWatch) de forma privada, sin pasar por el internet público.