AI Security & Compliance Engineer

Prueba Técnica

Prueba Técnica: AI Security & Compliance Engineer

Introducción

Esta prueba técnica está diseñada para evaluar sus conocimientos y habilidades en seguridad, privacidad y cumplimiento normativo aplicados a sistemas de IA. Se evaluarán tanto aspectos teóricos como prácticos, incluyendo:

- Conocimiento de marcos regulatorios relevantes para IA
- Capacidad de identificar y mitigar riesgos de seguridad en sistemas de IA
- Habilidad para diseñar controles y políticas de gobernanza
- Comprensión práctica de aspectos técnicos de seguridad en modelos de IA

Instrucciones

- Tiempo asignado: 90 minutos
- Complete todas las secciones
- Justifique sus respuestas con explicaciones claras y concisas
- Para los casos prácticos, incluya ejemplos de código o pseudocódigo cuando sea relevante

Contenido de la Prueba

Parte 1: Conocimientos Teóricos (30 puntos)

Explique los principales desafíos de seguridad específicos de los Large Language Models (LLMs) y cómo abordarlos. (10 puntos)

Describa los requerimientos clave del AI Act de la UE en relación con sistemas de IA de alto riesgo y cómo implementaría el cumplimiento técnico de estos. (10 puntos)

Enumere y explique tres técnicas para prevenir y detectar ataques de prompt injection en sistemas conversacionales de IA. (10 puntos)

Parte 2: Caso Práctico (40 puntos)

Escenario: Su empresa está desarrollando un asistente virtual de IA para el sector financiero que procesará información sensible de clientes.

Desarrolle:

Un framework de evaluación de riesgos que incluya:
- - Identificación de riesgos principales
- - Controles de mitigación
- - Métricas de monitoreo

(20 puntos)

Una propuesta técnica de arquitectura segura que incluya:
- - Diagrama de arquitectura
- - Controles de acceso y autenticación
- - Mecanismos de auditoría y logging
- - Protección de datos sensibles

(20 puntos)

Parte 3: Ejercicio Práctico (30 puntos)

Revise el siguiente fragmento de código de un sistema de IA y:

Identifique vulnerabilidades de seguridad
Proponga correcciones específicas
Agregue controles de validación y sanitización

```
def process_user_input(user_prompt):
# Conexión a la API del modelo
api_key = "sk_12345"
# Procesamiento directo del input
response = ai_model.generate(user_prompt)
# Almacenamiento de la interacción
db.store_interaction(user_prompt, response)
return response
```

Criterios de Evaluación

Conocimientos Teóricos (30%)

- Comprensión profunda de conceptos de seguridad en IA
- Conocimiento actualizado de regulaciones
- Claridad en explicaciones técnicas

Caso Práctico (40%)

- Exhaustividad en identificación de riesgos
- Viabilidad y efectividad de controles propuestos
- Claridad y completitud de la arquitectura

Ejercicio Práctico (30%)

- Identificación correcta de vulnerabilidades
- Calidad y seguridad de las soluciones propuestas
- Consideración de mejores prácticas de seguridad

Entrega

- Envíe sus respuestas en un documento PDF
- Incluya diagramas o código cuando sea necesario
- Nombre el archivo: "AISecurity_[SuNombre]_[Fecha]"

La evaluación considerará la profundidad técnica, claridad de comunicación y enfoque práctico de las soluciones propuestas.