

## 1. MLOps y AIOps : La IA como artefacto de software

- La IA deja de ser caja negra y se trata como componente de la cadena (Supply chain)
- **Unificación** : el modelo ML debe tener control de versiones, firma digital, SBOM
  - **AIOps en SRE** : Usar IA para digerir los logs y alertas (reducción de ruido y anomalías)  
Si la IA tiene permiso para ejecutar **playbooks** de remediación, se convierte en una superficie de ataque
  - **ML SecOps** : Aplicar seguridad al entrenamiento. Evitar datos manipulados y que el modelo no filtre secretos

## 2. LLMOpss y Guardrails : Comando al Copiloto

Cuando integramos LLM como copiloto, debemos limitar su capacidad de actuar

- **patrones de uso** : copilot sugiere código o test, chatbot que consulta a prometheus  
"Porque subio la latencia?"
- **Guardrails (Barandillas)** : Capas que validan que los LLM no ejecuten comandos  
Peligrosos, no filtre datos sensibles y respete esquemas JSON
- **OWASP LLM TOP** : Protegerse contra Prompt Injection, Excesiva agencia y Exfiltración de datos.

## 3. IA Protegiendo (y atacando) el pipeline

- \* **Código generado por IA = no confiable** : 40% de código generado puede ser inseguro, debe pasar por SAST/SCA
- \* **Usar IA para** : Generar políticas de seguridad y priorizar vulnerabilidades
- \* **Riesgo de "Vibe Coding"** : Aceptar código sugerido sin entenderlo o auditarlo, creando deuda técnica

## 4. Gobernanza y Riesgo :

- \* Tratar prompts como datos sensibles, aplicar minimización de datos para evitar fugar PII
- \* Usar Herramientas de auditoría para asegurar que el código generado no viole licencias
- \* Su filosofía es "menos puertas (gates), más barandillas (guardrails)"

## 5. Observabilidad de IA : no basta con monitorear el CPU del servidor, necesitamos nuevas métricas :

- **Drift** : ¿Los datos de entra han cambiado respecto a los de entrenamiento?
- **Calidad** : Alucinaciones, toxicidad, eficacia de las respuestas
- **Dashboard de salud del copiloto** : ¿Cuántas sugerencias acepta el equipo? ¿Cuántas introducen bug? ¿Cuántos prompts son bloqueados por los guardrails?