IACADEMY

DE FUNDAMENTOS A ARQUITECTURA DE IA

MÓDULO 09

IA para ciberseguridad

Especialización

iacedemy.com — 2026

MÓDULO 09

IA para ciberseguridad

Nivel: Especialización

Autor: Ricardo Gutierrez

Publicación: Mayo 2026

Plataforma: iacedemy.com

Pipeline SOC con 4 agentes

Este es el modulo mas critico del curso. Aqui un error de la IA no es un blog post malo: es un incidente que se cuela, un falso negativo que cuesta miles de euros.

```
 Alerta SIEM/EDR → Triage Agent → Investigation Agent → [HITL] → Reporting Agent
                                                ↕
                                  Threat Hunter (proactivo)
```

El SOC moderno recibe 10.000 alertas diarias. El 95% son falsos positivos. Un analista N1 revisa la alerta numero 200 del dia con menos atencion que la numero 20. Ahi se cuelan los incidentes reales.

Triage Agent

El filtro inicial. Recibe alertas crudas y las clasifica en 5 niveles: FP, BAJO, MEDIO, ALTO, CRITICO. La clave del diseno: si hay duda entre MEDIO y ALTO, siempre sube. Un falso negativo en seguridad es inaceptable.

Ejemplos de clasificacion:

- **FP:** Scanner desde IP conocida. Cierre automatico con justificacion.
- **BAJO:** Login fallido aislado.
- **ALTO:** Multiples login fallidos seguidos de exito. Posible credential stuffing.
- **ALTO:** Trafico DNS a dominio registrado hace 2 dias. Posible C2.
- **CRITICO:** Exfiltracion detectada. Escalar inmediatamente.

Investigation Agent

Toma las alertas ALTO y CRITICO. Genera timeline, identifica IOCs, mapea contra MITRE ATT&CK, determina alcance y recomienda contencion. **Recomendar, no ejecutar. Nunca ejecutar.**

Checkpoint HITL

Aqui entra el humano. El analista N2 valida el veredicto, confirma el alcance y autoriza la contencion. La IA preparo todo el trabajo. El humano toma la decision. Siempre.

Threat Hunter

No espera alertas: busca proactivamente. Genera hipotesis (login desde IP nueva despues de 3 intentos fallidos, proceso en directorio temporal con conexiones salientes) y propone queries para el SIEM.

Reporting Agent

Convierte la investigacion tecnica en 3 documentos:

- **Informe tecnico:** para el equipo SOC
- **Informe ejecutivo:** para el CISO (sin jerga, enfocado en impacto de negocio)
- **Ficha de evidencia:** para auditoria

GRC automatizado

Lo que a un consultor GRC le cuesta 3 dias, los agentes lo hacen en 10 minutos. Cuatro agentes GRC: Control Reviewer, Compliance Mapper, Evidence Collector, Audit Summary.

Compliance Mapper

Toma el inventario de controles del sistema y los mapea contra ENS, ISO 27001, NIS2 y DORA simultaneamente. Identifica controles que cubren multiples frameworks (eficiencia) y donde hay gaps.

Control Reviewer

Evaluacion ejemplo:

- TLS 1.3 inter-servicio: **CUMPLE** (ENS Art. 24)
- RLS en PostgreSQL: **CUMPLE** (ISO A.8.3)
- Backup cifrado: **PARCIAL** (backup existe pero sin cifrado en reposo)

- MFA: **NO CUMPLE** (solo en admin, falta en usuarios)
- Audit trail: **CUMPLE**

Evidence Collector

Lee Dockerfiles, configs de CI, settings de infraestructura. Para cada control genera una ficha con fuente de evidencia, verificacion, timestamp y hash SHA256. Evidencia trazable y auditable.

Audit Summary

5 controles evaluados: 3 cumplen, 1 parcial, 1 no cumple. Gap analysis con plan de accion, responsable y plazo para cada gap. Listo para el auditor externo.

> **Validacion obligatoria**
>
> Todo informe de compliance generado por IA necesita revision del CISO para validez tecnica, de legal si hay implicaciones regulatorias, y del auditor externo si es para certificacion. La IA genera borradores de calidad. La decision es humana.

Threat Intelligence con IA

Analisis de IOCs

IP sospechosa: buscar en AbuseIPDB, VirusTotal, OTX. Resultado: MALICIOSO, 15 reportes, asociado a botnet Mirai. Hash sospechoso: buscar en MalwareBazaar. Resultado: Emotet dropper, primera vista hace 48 horas.

Correlacion cruzada

La correlacion cruzada es donde la IA aporta mas valor. Un IOC que aparece en abuse.ch y en OTX tiene alta confianza. Si ademas afecta a tu stack tecnologico, es accionable. La IA lee feeds, correlaciona, y genera un informe semanal priorizado.

Generacion automatica de defensas

Basandose en el incidente investigado, la IA genera:

- Regla Sigma para el SIEM
- Query de Elasticsearch para busqueda retrospectiva
- IOCs en formato STIX 2.1 para compartir con otros SOCs

El analista valida la regla antes de anadirla al SIEM.

Pipeline n8n de monitoring CTI

```
 Trigger cada 6 horas → Fetch abuse.ch, MalwareBazaar, OTX
→ Correlacion con stack del cliente
→ Si IOC relevante: alerta Slack + ticket automatico
→ Resumen semanal por email al equipo de seguridad
```

Las 7 reglas de oro

1. **NUNCA contencion sin aprobacion humana.** La IA recomienda, el humano ejecuta.
2. **NUNCA decidir sin validacion.** El output de la IA es un borrador, no una decision.
3. **NUNCA acceso directo a produccion.** La IA opera sobre replicas o logs exportados.
4. **NUNCA IOCs inventados.** Una alucinacion en seguridad genera fatiga de alertas y desconfianza.
5. **NUNCA comunicar a terceros sin Legal y CISO.** Reguladores, clientes, prensa: aprobacion obligatoria.
6. **NUNCA almacenar evidencia fuera de sistemas controlados.** Cadena de custodia siempre.
7. **NUNCA operar sin audit trail.** Toda accion registrada con timestamp, agente, input, output.

Riesgos especificos de IA en seguridad

Riesgo	Descripcion	Mitigacion
		Input sanitization

Riesgo	Descripcion	Mitigacion
Prompt injection	Atacante incluye instrucciones en ticket que manipulan al agente	
Data leakage	IA expone datos de un cliente al analizar alertas de otro	Aislamiento de contexto por tenant
Hallucinated IOCs	IA inventa un hash, se bloquea en firewall, causa falso positivo	Restriccion anti-alucinacion agresiva

AI-assisted pentesting

La fase de recon es la mas mecanica y la que mas se beneficia de IA. Subdominios, tecnologias, endpoints, emails corporativos, leaks en GitHub. La IA sistematiza. El pentester interpreta y decide los vectores.

La IA puede revisar un endpoint de login buscando Broken Access Control, Injection, Authentication Failures. Para cada hallazgo: severidad, CWE-ID, PoC conceptual y fix con codigo. Pero la validacion manual es obligatoria. La IA genera falsos positivos que un pentester descarta en segundos, y puede perder vulnerabilidades logicas que requieren contexto de negocio.

Ejercicio practico

Ejercicio M09: Pipeline SOC simplificado

1. Crea un Triage Agent como slash command que clasifique alertas en 5 niveles.
2. Dale 5 alertas simuladas y verifica que clasifica correctamente.
3. Crea un Investigation Agent que analice las alertas ALTO/CRITICO.
4. Anade el checkpoint HITL: el agente NO ejecuta, solo recomienda.
5. Crea un Reporting Agent que genere informe tecnico + ejecutivo.
6. Configura un eval dataset con 20 alertas anotadas para medir accuracy del Triage.

> **Bonus:** Configura un workflow n8n que busque IOCs en abuse.ch cada 6 horas y notifique en Slack si hay match con tu stack.

Conclusiones clave

> **Key takeaways del M09**
>
> 1. Pipeline SOC con 4 agentes: Triage, Investigation, Threat Hunter, Reporting. Siempre con HITL.
> 2. GRC automatizado mapea controles contra 4 frameworks simultaneamente en minutos.
> 3. Threat intel con correlacion cruzada de feeds aporta confianza y accionabilidad.
> 4. Las 7 reglas de oro son no negociables. El humano decide. Siempre.
> 5. IA en pentesting: excelente para recon, limitada en vulnerabilidades logicas.
> 6. Los riesgos (injection, leakage, alucinaciones) tienen mitigaciones concretas.

La IA en seguridad es la mas poderosa y la mas peligrosa

En el M10 cambiamos de vertical: IA para negocio y productividad. Market research, contenido a escala, ventas asistidas y ROI medible.

Ir al Modulo 10

IACADEMY

iacedemy.com

De fundamentos a arquitectura de IA.
12 módulos prácticos. 24 recursos descargables.
Quizzes con certificado. Vídeos profesionales.

Empieza gratis en iacedemy.com/free