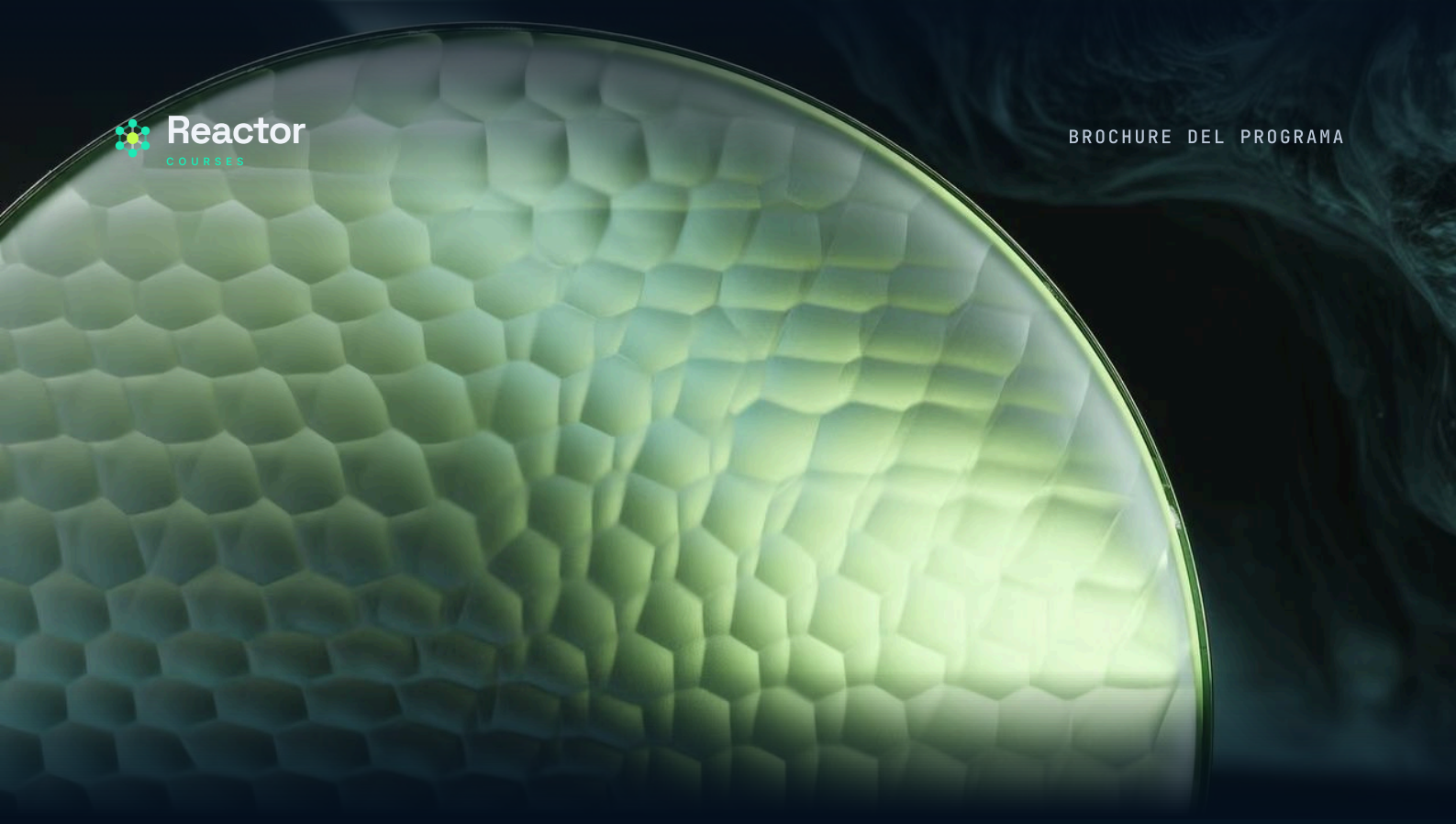


Reactor
COURSES

BROCHURE DEL PROGRAMA

ARQUITECTURA E INGENIERÍA · INTELIGENCIA ARTIFICIAL

# Arquitectura de seguridad y gobierno de IA

## Atacar, medir y gobernar sistemas con IA

Asegura y gobierna sistemas que usan modelos de IA con las dos manos: atacando el propio sistema y midiendo cada defensa, y armando el inventario, la evaluación de impacto y el mapa de controles que pide una revisión de riesgo.

**Lo que vas a lograr** Aprendes a atacar tu propio sistema con IA para encontrar sus fallas, a medir cada defensa y a documentar los riesgos como lo pide una revisión formal.

- OWASP LLM Top 10
- Red teaming
- Defensas medidas
- Datos personales
- NIST AI RMF e ISO 42001
- Incidentes de IA

| horas en vivo | sesiones | semanas | riesgos del OWASP LLM Top 10 trabajados |
|---|---|---|---|
| 60 | 20 | 10 | 10 |

**INICIO**
Por confirmar

**PRECIO**
**US$360** pago único · USD

**MODALIDAD**
Clases en vivo por Zoom · Avanzado

reactorcourses.com/programas/ai-security-governance-architecture

EL PROGRAMA

# Un sistema con IA no se asegura con un filtro. Se ataca, se mide y se gobierna.

Asegura y gobierna sistemas que usan modelos de IA con las dos manos: atacando el propio sistema y midiendo cada defensa, y armando el inventario, la evaluación de impacto y el mapa de controles que pide una revisión de riesgo. Un programa de **60 horas** para ingenieros de seguridad que reciben sistemas con IA para revisar, ingenieros de IA que tienen que pasar esa revisión y responsables de riesgo que deben armar un programa de gobierno.

| | | | |
|---|---|---|---|
| Duración | 60 horas · 20 sesiones | Calendario | 10 semanas · 2 sesiones por semana de 3 h |
| Modalidad | Clases en vivo por Zoom | Grabaciones | Disponibles por 12 meses |
| Credenciales | 3 certificados con insignia de Reactor Courses | Nivel | Avanzado |
| Inicio | Por confirmar | Horario | Lunes y miércoles, 19:00–22:00 (GMT-5) |
| Precio | US$360 · pago único | | |

Los sistemas con modelos de lenguaje reciben contenido no confiable por todos lados —el usuario, los documentos, las respuestas de sus herramientas— y toman decisiones con datos de personas. Este programa enseña a atacar el propio sistema, a poner defensas cuyo efecto se mide, y a armar el gobierno que pide un comité de riesgo: inventario, evaluación de impacto, controles y respuesta a incidentes.

01

### Atacar y medir

Amenazas, red teaming, defensas y privacidad

Modelado de amenazas con OWASP y MITRE ATLAS, campañas de red teaming con herramientas abiertas contra un sistema propio, y defensas en capas con su tasa de bloqueo y sus falsos positivos medidos sobre el mismo conjunto de ataques. Los datos personales se protegen en lo que entra al modelo, en lo que sale y en lo que queda registrado.

- Inyección directa, indirecta y abuso de herramientas
- Exfiltración, envenenamiento y respuestas que dañan sin ataque
- Defensas en capas con tasa de bloqueo y falsos positivos
- Datos personales y cadena de suministro de modelos

02

### Gobernar

Inventario, impacto, marcos, normativa e incidentes

Inventario de sistemas con IA, incluidos los que nadie declaró; evaluación de impacto proporcional al riesgo; mapeo de controles al NIST AI RMF y a la estructura de ISO/IEC 42001 sin afirmar conformidad; el EU AI Act para empresas de la región que operan con Europa; la normativa de datos personales traducida a controles; y la respuesta a un incidente de IA.

- Inventario y clasificación de riesgo
- Evaluación de impacto con medición de sesgo
- NIST AI RMF, ISO/IEC 42001 y EU AI Act con fechas vigentes
- Respuesta a incidentes de IA con reconstrucción

> Ninguna defensa contra la inyección de instrucciones es completa. La pregunta útil no es si bloquea, sino cuánto bloquea, cuánto molesta al cliente legítimo y qué riesgo queda.

PARA QUIÉN ES

## ¿Es el programa indicado para ti?

Está pensado para quienes protegen o auditan sistemas con IA y pueden leer y modificar código Python. En los laboratorios atacas y defiendes un sistema real, paso a paso.

**Es para ti si**

- ✓ Ingenieros de seguridad que reciben sistemas con IA para revisar.
- ✓ Ingenieros de IA que tienen que pasar una revisión de seguridad o de riesgo.
- ✓ Responsables de riesgo tecnológico que deben armar un programa de gobierno de IA.
- ✓ Arquitectos que diseñan asistentes con acceso a datos de clientes.
- ✓ Equipos de cumplimiento que necesitan traducir normas a controles verificables.

**Te conviene otro programa**

- – Quien no pueda leer y modificar código Python: los laboratorios atacan y defienden sistemas reales.
- – Quien busque preparar una certificación de auditor de ISO/IEC 42001 u otra: no se prepara ninguna.
- – Abogados que buscan derecho regulatorio de IA: aquí la norma se traduce a controles, no se interpreta jurídicamente.
- – Quien espere una defensa completa contra la inyección de instrucciones: se enseña a medir y declarar el riesgo que queda.
- – Quien ya cursó un programa de agentes o de plataforma de IA: cerca de un tercio de este programa —modelado de amenazas, inyección indirecta, abuso de herramientas, defensas, privacidad, inventario de riesgo e incidentes— toca temas que esos programas tratan al construir. Lee en las preguntas frecuentes la que trata ese caso.
- – Quien no pueda dedicar entre 9 y 11 horas por semana durante 10 semanas.

EL RESULTADO

## Terminas con la revisión de seguridad y gobierno completa de un sistema con IA, defendida ante un comité.

Cada laboratorio deja una pieza de la revisión del asistente del caso. Juntas forman lo que defiendes al final ante un comité de riesgo simulado.

01 **Modelo de amenazas** Diagrama de flujo de datos, quince amenazas priorizadas y ocho casos de prueba adversarial derivados.

02 **Línea base de ataque** Campaña automatizada con su registro de evidencia y tasa de éxito por categoría.

03 **Inyecciones documentadas** Inyección directa en varios turnos e inyección indirecta por documentos y por la herramienta, rastreadas hasta el fragmento responsable.

04 **Abuso y fuga demostrados** Consulta de la cuenta de otro cliente, consumo desmedido y exfiltración de datos documentados con trazas y costo.

05 **Envenenamiento detectado** Reglamento falso que cambia respuestas, detectado comparando procedencia y huellas de los documentos.

06 **Defensas con números** Tres capas de defensa con tasa de bloqueo y falsos positivos sobre el mismo conjunto de ataques y preguntas legítimas.

07 **Datos personales protegidos** Enmascarado antes del modelo y en registros, con lo que se escapa medido, y una solicitud de borrado atendida en todo el sistema.

08 **Modelo firmado e inventariado** Modelo local firmado, inventario de componentes de IA y un despliegue que rechaza una firma inválida.

09 **Respuestas dañinas reducidas** Conjunto propio de preguntas que inducen asesoría indebida, con la cifra de respuestas dañinas medida y reducida.

10 **Inventario y evaluación de impacto** Inventario con usos no declarados detectados, clasificación de riesgo y evaluación de impacto con medición de sesgo.

11 **Controles mapeados sin promesas** Controles con evidencia mapeados al NIST AI RMF y a la estructura de ISO/IEC 42001, con las brechas declaradas.

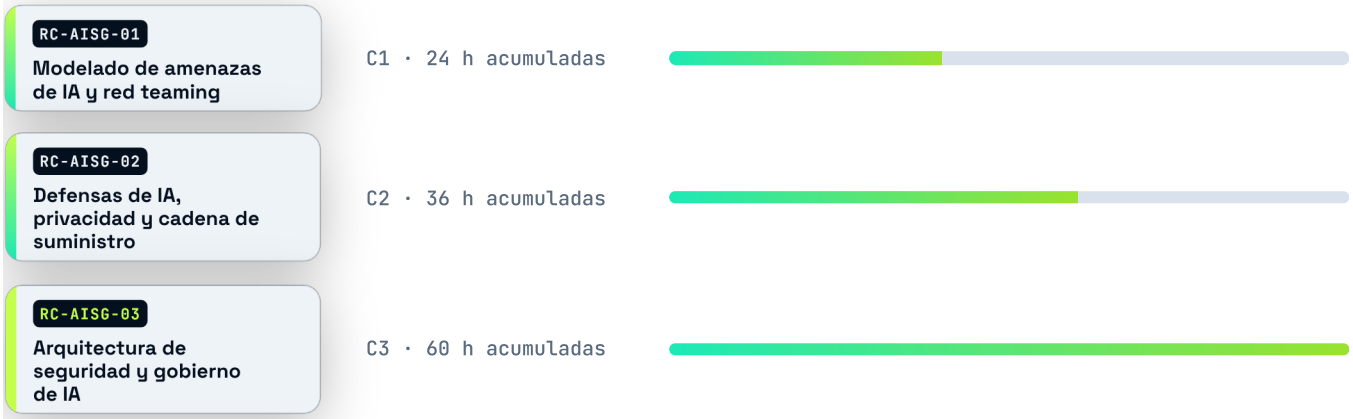
12 **Incidente resuelto** Fuga simulada contenida, reconstruida con la trazabilidad y verificada con la regresión adversarial.

---

RUTA DE CERTIFICACIÓN

## 3 credenciales progresivas, no una al final.

Cada etapa cierra un bloque completo: primero amenazas y ataques, después defensas, privacidad y cadena de suministro, y al final gobierno y respuesta. Son credenciales de Reactor Courses: no son certificaciones de auditor ni acreditan conformidad con ningún marco.




CERTIFICACIÓN INTEGRAL

**AI Security & Governance Architecture — Atacar, medir y gobernar sistemas con IA (60 horas)**

Requiere entregar la revisión de seguridad y gobierno del asistente con la campaña, las defensas medidas, la matriz de controles y el ejercicio de incidente. Los certificados son emitidos por Reactor Courses y acreditan las horas y competencias del programa.

Cada certificado de Reactor Courses viene con su insignia, aparte de cualquier certificación oficial.

Cualquiera puede comprobar un certificado de Reactor Courses con su código en reactorcourses.com/verificar.

TEMARIO

## Las 20 sesiones, abiertas y sin letra chica.

60 horas repartidas en 5 módulos. Así sabes qué verás, qué practicarás y qué entregarás en cada una de las 10 semanas.

MÓDULO 1 **AMENAZAS Y MÉTODO** — 9 h · 3 sesiones

01 **La superficie de ataque de un sistema con IA**

el caso: asistente de una caja de ahorro con RAG sobre reglamentos y una herramienta de consulta de cuentas · OWASP Top 10 for LLM Applications 2025 como mapa · MITRE ATLAS como vocabulario de tácticas y técnicas · qué cambia frente a la seguridad de aplicaciones clásica · el modelo no es el único activo: datos, corpus, herramientas y registros · reglas del programa: solo se atacan sistemas propios del curso

LABORATORIO: Levantar el asistente del caso y recorrer su superficie con el OWASP LLM Top 10, entregando el inventario de activos —modelo, corpus, herramientas, registros y usuarios— con el riesgo del Top 10 que aplica a cada uno

02 **Modelado de amenazas de un sistema con IA**

diagrama de flujo de datos con límites de confianza del modelo · dónde entra contenido no confiable: usuario, documentos y respuestas de herramientas · amenazas por componente con OWASP y ATLAS · priorización por impacto en el cliente y probabilidad · supuestos de seguridad que el equipo nunca escribió · del modelo de amenazas a la lista de pruebas adversariales

`LABORATORIO:` Modelar las amenazas del asistente con un diagrama de flujo de datos y técnicas de MITRE ATLAS, priorizar quince amenazas y convertir las ocho principales en casos de prueba adversarial que se ejecutarán en el resto del programa

03 **Método de red teaming**

alcance, reglas de compromiso y criterio de corte · campaña manual frente a campaña automatizada · garak, PyRIT y promptfoo: qué hace bien cada uno · registro de campaña: intento, resultado y evidencia · tasa de éxito de ataque como métrica · reproducibilidad: semilla, versión de modelo e instrucciones

`LABORATORIO:` Ejecutar una primera campaña automatizada con garak contra el asistente, registrar cada intento con su evidencia y calcular la tasa de éxito de ataque por categoría como línea base del programa

## `MÓDULO 2` ATAQUES

`15 h · 5 sesiones`

04 **Inyección directa y evasión de instrucciones**

inyección directa y jailbreak: técnicas vigentes · fuga de las instrucciones de sistema (LLM07) · ataques en varios turnos · codificaciones, idiomas y formatos para evadir filtros · ataques en español frente a ataques en inglés · medir el éxito sin depender del juicio de quien ataca

`LABORATORIO:` Lanzar con PyRIT una campaña de inyección directa en varios turnos y en español contra el asistente, logrando y documentando al menos una fuga de instrucciones, y medir la tasa de éxito por técnica

05 **Inyección indirecta por el corpus y por herramientas**

instrucciones escondidas en documentos indexados · respuestas de herramientas como vector de ataque · debilidades de vectores y embeddings (LLM08) · inyección que se activa con una pregunta inocente · el usuario como víctima, no como atacante · rastrear qué fragmento disparó la acción

`LABORATORIO:` Plantar tres documentos con instrucciones ocultas en el corpus y una respuesta manipulada de la herramienta, demostrar que el asistente las ejecuta ante preguntas legítimas de otro usuario y rastrear con las trazas el fragmento responsable

06 **Abuso de herramientas y agencia excesiva**

agencia excesiva (LLM06): permisos, funciones y autonomía · parámetros de herramienta controlados por el atacante · confusión de identidad: el asistente actúa con permisos que el usuario no tiene · manejo inseguro de la salida del modelo (LLM05) · acciones irreversibles sin confirmación · consumo sin límite (LLM10) como ataque de costo

`LABORATORIO:` Hacer que el asistente consulte la cuenta de otro cliente mediante inyección en los parámetros de la herramienta y provocar un consumo desmedido de tokens, documentando ambos con la traza y el costo generado

07 **Exfiltración y fuga de información**

divulgación de información sensible (LLM02) · exfiltración por enlaces e imágenes en la respuesta · datos de un usuario en la respuesta de otro · memoria y caché como fuente de fuga · registros que guardan lo que no deberían · extraer el corpus pregunta a pregunta

`LABORATORIO:` Exfiltrar datos del corpus restringido mediante un enlace en la respuesta y mediante preguntas encadenadas, y encontrar en los registros del asistente datos personales que no deberían estar ahí

08 **Envenenamiento de datos y del corpus**

envenenamiento de datos y de modelos (LLM04) · documentos falsos con autoridad aparente · envenenamiento de ejemplos de ajuste y de evaluación · desinformación con apariencia de fuente (LLM09) · procedencia del documento como defensa · detección de cambios anómalos en el corpus

`LABORATORIO:` Introducir en el corpus un reglamento falso con formato oficial y demostrar que cambia las respuestas sobre tasas de interés, luego detectar la alteración comparando procedencia y huellas de los documentos

## MÓDULO 3 DEFENSAS, PRIVACIDAD Y CADENA DE SUMINISTRO — 12 h · 4 sesiones

### 09 Defensas en capas medidas

filtros de entrada, delimitación del contenido no confiable y filtros de salida · NeMo Guardrails como capa programable · mínimo privilegio en herramientas y confirmación humana · tasa de bloqueo y falsos positivos sobre el mismo conjunto de ataques · costo en latencia y en experiencia del cliente legítimo · riesgo residual declarado

**LABORATORIO:** Implementar tres capas de defensa en el asistente —una de ellas con NeMo Guardrails— y reportar para cada una la tasa de bloqueo y la tasa de falsos positivos sobre el mismo conjunto de ataques de las sesiones anteriores y sobre cien preguntas legítimas

### 10 Datos personales en prompts, respuestas y registros

detección y enmascarado de datos personales con Presidio · qué se guarda en los registros y por cuánto tiempo · minimización: no enviar al modelo lo que no necesita · datos personales en el corpus indexado · transferencia de datos a un proveedor de modelos en otro país · derechos del titular frente a un sistema con IA

**LABORATORIO:** Enmascarar con Presidio los datos personales antes del modelo y en los registros, medir lo que se escapa sobre doscientos mensajes de prueba en español, y atender una solicitud de borrado de un cliente en corpus, registros y caché

### 11 Cadena de suministro de modelos y datos

riesgos de la cadena de suministro (LLM03) · modelos de pesos abiertos: procedencia y formatos seguros · firma de modelos con model-transparency de Sigstore · inventario de componentes de IA con CycloneDX · licencias de modelos y de conjuntos de datos · proveedor de modelo que cambia su versión sin aviso

**LABORATORIO:** Firmar y verificar el modelo local del caso con model-transparency, generar el inventario de componentes de IA en CycloneDX y demostrar que el despliegue rechaza un modelo con firma inválida

### 12 Respuestas que dañan sin que haya un ataque

desinformación (LLM09) en un dominio regulado · asesoría financiera que el asistente no debe dar · límites del dominio y un rechazo que igual ayuda · citar la fuente o abstenerse · medir respuestas dañinas con un conjunto propio en español · derivación a una persona cuando corresponde

**LABORATORIO:** Construir con promptfoo un conjunto de cincuenta preguntas que inducen asesoría indebida o tasas inventadas, medir cuántas respuestas dañarían a un cliente, y reducir esa cifra con límites de dominio y abstención sin bloquear las preguntas legítimas

## MÓDULO 4 GOBIERNO — 15 h · 5 sesiones

### 13 Inventario y clasificación de riesgo

inventario de sistemas con IA, incluidos los que nadie declaró · criterios de clasificación de riesgo · categorías de riesgo del EU AI Act como referencia · dueño, propósito, datos y proveedor de cada sistema · uso de IA por empleados fuera de control · actualización del inventario como proceso

**LABORATORIO:** Construir el inventario de sistemas con IA de la caja de ahorro a partir de entrevistas simuladas y registros de red, clasificar cada sistema por riesgo con criterios escritos y detectar dos usos no declarados

### 14 Evaluación de impacto proporcional al riesgo

evaluación de impacto: cuándo y con qué profundidad · impacto en personas: sesgo, exclusión y error con daño · evaluación de sesgo con datos del caso · supervisión humana proporcional al riesgo · condiciones de salida a producción · registro de decisión firmado por el responsable

**LABORATORIO:** Redactar la evaluación de impacto del asistente con una medición de sesgo en respuestas sobre productos de crédito por grupo de clientes, y fijar las condiciones medibles que el sistema debe cumplir para salir a producción

15 **NIST AI RMF y el perfil de IA generativa**

funciones Govern, Map, Measure y Manage · perfil de IA generativa NIST AI 600-1 · revisión del marco en curso y cómo seguirla · mapear controles técnicos a categorías del marco · evidencia para cada control · brechas declaradas en lugar de cumplimiento declarado

LABORATORIO: Mapear los controles implementados en el programa a las funciones del NIST AI RMF y a los riesgos del perfil de IA generativa, con la evidencia de cada uno y las brechas declaradas, sin afirmar conformidad

16 **ISO/IEC 42001 y EU AI Act para empresas latinoamericanas**

ISO/IEC 42001:2023 como estructura de un sistema de gestión de IA · políticas, roles, objetivos y mejora continua · cuándo el EU AI Act alcanza a una empresa latinoamericana · fechas de aplicación vigentes y su revisión · proveedor y desplegador: obligaciones distintas · certificar frente a mapear: qué hace este programa

LABORATORIO: Diseñar la política de uso de IA y la estructura mínima de gestión de la caja de ahorro siguiendo la estructura de ISO/IEC 42001, y determinar con fechas si el asistente quedaría alcanzado por el EU AI Act en tres escenarios de negocio

17 **Normativa de datos personales aplicada a la IA**

Ley 29733 de Perú y su reglamento aplicados a un asistente · base legal del tratamiento y consentimiento · transferencia internacional de datos al proveedor del modelo · comparación con la ley de otro país de la región presente en el grupo · registro de bancos de datos y medidas de seguridad · traducir la norma a controles, no interpretarla jurídicamente

LABORATORIO: Traducir los requisitos de la Ley 29733 y su reglamento aplicables al asistente en una lista de controles verificables, compararlos con los de una ley de otro país de la región y marcar qué controles del programa ya los cubren

## MÓDULO 5 RESPUESTA Y PROYECTO

9 h · 3 sesiones

18 **Incidentes de IA**

qué es un incidente de IA y cómo se clasifica · contención: apagar una herramienta, cambiar el modelo, sacar un documento · reconstrucción con la trazabilidad · notificación a afectados y a la autoridad cuando corresponde · corregir y verificar con la regresión adversarial · aprendizaje que vuelve al modelo de amenazas

LABORATORIO: Responder a un incidente simulado de fuga de datos de un cliente por inyección indirecta: contenerlo, reconstruir qué pasó con la trazabilidad, evaluar si corresponde notificar y verificar la corrección con la regresión adversarial

19 **Revisión integral de seguridad y gobierno**

consolidar modelo de amenazas, campaña y defensas · métricas antes y después sobre el mismo conjunto de ataques · matriz de controles con evidencia · riesgo residual aceptado y quién lo firma · revisión cruzada entre equipos · preparación de la defensa

LABORATORIO: Ejecutar la revisión de seguridad y gobierno del sistema de otro equipo con su campaña de red teaming, reportar las métricas antes y después y emitir los hallazgos priorizados que ese equipo deberá corregir

20 **Defensa ante el comité**

presentación a un comité de riesgo · resultados de la campaña y defensas medidas · inventario, evaluación de impacto y mapeo a marcos · ejercicio de incidente y lecciones · riesgo residual declarado · preguntas cruzadas del grupo

LABORATORIO: Defender ante un comité simulado la revisión de seguridad y gobierno del asistente con la campaña, las defensas medidas antes y después, la matriz de controles y el riesgo residual declarado

LO QUE VAS A USAR

## Herramientas abiertas y marcos públicos, sin aval de ninguno de sus autores.

### Taxonomías

- OWASP Top 10 for LLM Applications 2025
- MITRE ATLAS

### Red teaming

- garak
- PyRIT
- promptfoo

### Defensas y privacidad

- NeMo Guardrails
- Presidio

### Cadena de suministro

- model-transparency de Sigstore
- CycloneDX

### Marcos de gobierno

- NIST AI RMF 1.0
- NIST AI 600-1
- ISO/IEC 42001:2023
- EU AI Act

### Normativa de datos

- Ley 29733 de Perú y su reglamento
- Una ley de datos personales de la región

Los logos son marcas de sus respectivos dueños. Se muestran solo para identificar las herramientas que se usan en el programa y no implican aval, patrocinio ni afiliación.

---

CÓMO FUNCIONA

## Clases en vivo, un sistema propio para atacar y cada defensa medida con el mismo conjunto de ataques.

01 **Clases en vivo por Zoom** 20 sesiones de 3 horas, 2 veces por semana. Se ataca y se defiende en vivo, y se discuten los resultados con el grupo.

02 **Todo queda grabado** Si te pierdes una sesión o vives en otro huso horario, la grabación queda disponible. Acceso por 12 meses.

03 **Mitad técnica, mitad gobierno** Doce sesiones de amenazas, ataques y defensas, y ocho de gobierno, normativa e incidentes, unidas por el mismo sistema.

04 **Solo sistemas propios** Se atacan únicamente los sistemas construidos para el curso. Nunca servicios de terceros.

05 **Bloqueo y falsos positivos, juntos** Ninguna defensa se presenta solo con lo que bloquea: se reporta también cuántas preguntas legítimas rechaza.

06 **Normas con fecha y fuente** Todo dato normativo se enseña con su fuente y su fecha, porque la regulación de IA cambia mientras se dicta el programa.

### Horarios · Primer grupo

| | |
|---|---|
| Lunes | 19:00 - 22:00 |
| Miércoles | 19:00 - 22:00 |

Horario referido a GMT-5 (Perú, Colombia, Ecuador y Panamá; en Ciudad de México es una hora menos).

Los días y el horario son la propuesta para este grupo y se confirman junto con la fecha de inicio, antes de abrir la matrícula. El calendario definitivo se entrega al confirmar la matrícula.

REQUISITOS

# Lo que necesitas traer a la sesión 01.

No hay nivelación previa. La sesión 01 levanta el asistente del caso y empieza a buscarle problemas.

### Programación

- ✓ Python: leer y modificar un servicio
- ✓ HTTP y JSON
- ✓ Git

### IA

- ✓ Haber usado la API de un modelo de lenguaje
- ✓ Deseable: haber construido un RAG o un agente simple

### Seguridad

- ✓ OWASP Top 10 web a nivel de conceptos
- ✓ Nociones de autenticación y autorización

### Cuentas

- ✓ Cuenta de un proveedor de modelos con límite de gasto, o un modelo abierto local
- ✓ Docker a nivel de uso

### Tiempo

- ✓ Seis horas de clase por semana
- ✓ Entre tres y cinco horas de laboratorio
- ✓ 10 semanas de compromiso

**Costos de herramientas que asumes tú** `US$5 a 25 aprox.`

Estos costos NO están incluidos en el precio del programa. Las herramientas de red teaming y de defensa son de código abierto.

- — Consumo de la API de un modelo comercial para las campañas de ataque y las mediciones: del orden de US$5 a US$25 en todo el programa con un modelo pequeño
- — Alternativa sin costo de API: todos los laboratorios de ataque y defensa se pueden correr con un modelo abierto local si tu equipo tiene 16 GB de memoria
- — Se configura un límite de gasto en la cuenta del proveedor de modelos en la sesión 01

PRECIO

## 60 horas para revisar un sistema con IA como lo haría un comité de riesgo exigente.

**US$360**

Pago único · USD · Incluye impuestos

Todavía no hay fecha de inicio. Escríbenos y te avisamos cuando la confirmemos: no cobramos nada antes de abrir la matrícula.

### Incluye

- ✓ 20 sesiones en vivo por Zoom (60 horas lectivas)
- ✓ Grabación de todas las clases con acceso por 12 meses
- ✓ El asistente del caso listo para atacar, con su corpus, su herramienta y sus registros
- ✓ Conjunto de ataques en español y de preguntas legítimas para medir defensas
- ✓ Plantillas de modelo de amenazas, inventario, evaluación de impacto y matriz de controles
- ✓ Los 3 certificados progresivos
- ✓ Canal privado del grupo para dudas entre sesiones
- ✓ Actualizaciones del material por 12 meses

### No incluye

- – El consumo de la API de modelos para las campañas (aprox. US$5 a 25 en todo el programa, o cero con un modelo local)
- – Asesoría legal: la normativa se traduce a controles, no se interpreta jurídicamente
- – Certificaciones de auditor o de conformidad de ISO/IEC 42001 o de cualquier otro marco

### Métodos de pago

Tarjeta de crédito · Tarjeta de débito internacional · PayPal · Apple Pay · Google Pay

**Garantía de 14 días.** Si dentro de los primeros 14 días del programa decides que no es para ti, te devolvemos el 100 % de lo pagado. No pedimos justificación.

Paddle actúa como comerciante registrado (merchant of record): emite el comprobante y gestiona los impuestos de cada país.

PREGUNTAS FRECUENTES

## Preguntas frecuentes sobre el programa.

### ¿Las clases son en vivo o grabadas?

En vivo por Zoom: 20 sesiones de 3 horas, 2 veces por semana. Todas quedan grabadas y disponibles por 12 meses.

### ¿Cuándo empiezan las clases?

La fecha de inicio y el calendario definitivo todavía no están confirmados: se publican en esta misma ficha antes de que se abra la matrícula. Los días propuestos son Lunes y miércoles en horario de GMT-5 (Perú, Colombia, Ecuador y Panamá; en Ciudad de México es una hora menos). Si quieres que te avisemos cuando se confirmen, escribe a soporte@reactorcourses.com.

### ¿Qué nivel necesito?

Leer y modificar código Python, haber usado la API de un modelo de lenguaje y conocer las bases de la seguridad de aplicaciones. Es deseable haber construido un RAG o un agente simple.

### ¿Cuánto voy a gastar en modelos?

Con un modelo comercial pequeño, unos US$5 a 25 en todo el programa, sobre todo por las campañas automatizadas. Todos los laboratorios de ataque y defensa se pueden hacer con un modelo abierto local sin costo si tu equipo tiene 16 GB de memoria.

### Ya cursé un programa de agentes o de plataforma de IA. ¿Me sirve este?

Sí, y conviene que sepas qué se repite. Cerca de un tercio de las sesiones toca temas que esos programas tratan al construir: modelado de amenazas, inyección indirecta, abuso de herramientas, defensas en capas, privacidad, inventario de riesgo e incidentes. Aquí el centro es otro: atacar el sistema, medir cada defensa y armar el gobierno, con otros laboratorios. Lo nuevo serían la campaña de red teaming, exfiltración, envenenamiento, cadena de suministro, evaluación de impacto, mapeo a marcos y normativa. No hay descuento ni convalidación entre programas.

### ¿Me prepara para certificarme en ISO/IEC 42001?

No. Se usa la estructura de ISO/IEC 42001 para diseñar un sistema de gestión de IA y se mapean controles, pero el programa no prepara para ninguna certificación de auditor ni ayuda a obtener una certificación de conformidad. Tampoco afirma que el resultado cumpla con el NIST AI RMF ni con el EU AI Act: se mapea y se declaran las brechas.

### ¿Qué se ataca en los laboratorios?

Solo el asistente construido para el curso, en tu propio entorno. Nunca servicios de terceros. Las herramientas de red teaming se usan con alcance y criterio de corte definidos en la sesión 03.

### ¿Se ve el EU AI Act si mi empresa no está en Europa?

Sí, en lo que alcanza a empresas latinoamericanas que venden u operan sistemas con IA en la Unión Europea. Se trabaja con las fechas de aplicación vigentes al momento de dictar el programa, que el material cita con su fuente porque ya cambiaron una vez.

### ¿Qué pasa si falto a una clase?

La grabación queda disponible y cada laboratorio tiene su punto de partida en el repositorio. La revisión se construye por partes: faltar a varias sesiones seguidas sin recuperarlas te deja sin piezas para la defensa final.

### ¿Los certificados tienen validez oficial?

No. Son certificados de Reactor Courses que acreditan las horas y competencias del programa. No son títulos universitarios, certificaciones de auditor ni acreditaciones de conformidad.

### ¿Cuánto tiempo debo dedicarle por semana?

Seis horas de clase más entre tres y cinco horas de laboratorio. Considera entre 9 y 11 horas semanales durante 10 semanas.

### ¿Qué métodos de pago aceptan?

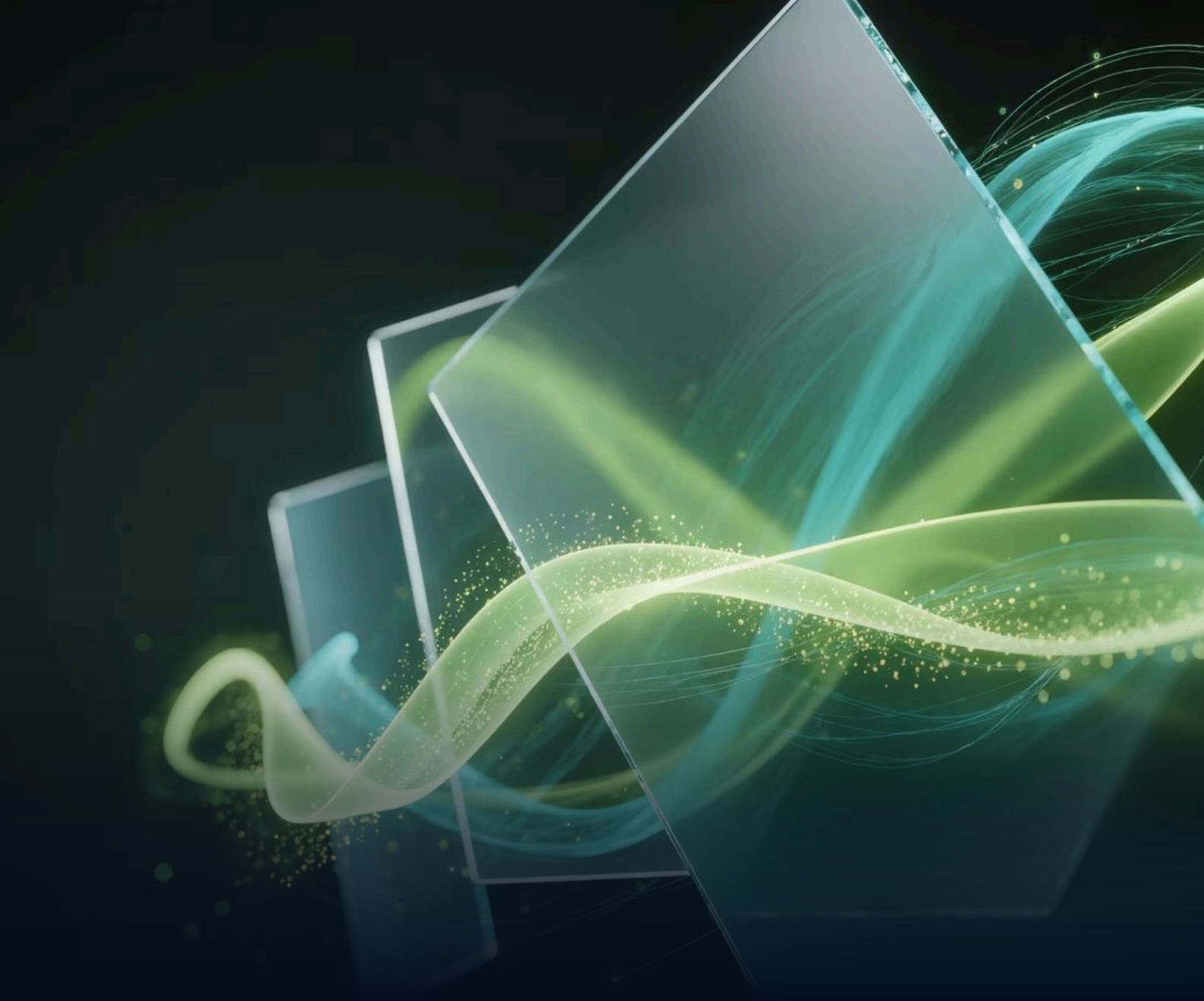
El pago lo procesa Paddle y acepta Tarjeta de crédito, Tarjeta de débito internacional, PayPal, Apple Pay, Google Pay. El cobro es en dólares. Importante: no procesamos Yape, Plin, OXXO, Mercado Pago ni transferencias locales. Si tu banco rechaza compras internacionales, habilítalas antes de pagar o usa PayPal.

### ¿Emiten factura o comprobante?

Sí. Paddle actúa como comerciante registrado (merchant of record): emite el comprobante y gestiona los impuestos de cada país. Lo recibirás por correo automáticamente al completar el pago.

### ¿Cuál es la política de reembolso?

Tienes 14 días desde el inicio del programa para pedir el reembolso completo escribiendo a soporte@reactorcourses.com. Publicamos el temario completo y los requisitos justamente para que decidas informado.



Reactor
COURSES

Ingeniería de software en la era de los agentes

## ¿Te quedó alguna duda?

Escríbenos a info@reactorcourses.com. La ficha al día, con el temario interactivo y la inscripción, está en reactorcourses.com/programas/ai-security-governance-architecture.

`reactorcourses.com/programas/ai-security-governance-architecture`

---

Información vigente al 23 de noviembre de 2026. El precio, las fechas y el temario pueden cambiar: la versión al día es la de la web.

Reactor Courses no está afiliado a Microsoft, Amazon Web Services, OpenAI, Anthropic ni Google. Los nombres de productos se mencionan únicamente con fines descriptivos del contenido académico.