

4 semanas | En vivo + material grabado · 15 cupos | Claude Code o Codex | Capacitación de seguridad informática

Blindá tu bot de WhatsApp, tus agentes y tus automatizaciones

Probás vos mismo, en un laboratorio propio, cómo se rompe un bot inmobiliario con IA y lo rediseñás para que no filtre datos ni te haga perder plata.

PARA QUIÉN

Corredores, programadores e inmobiliarias que ya tienen o van a poner un bot de WhatsApp, un agente o flujos en n8n, Make o Apps Script.

CÓMO TRABAJAN LOS AGENTES DE IA

Un agente de IA (Claude Code o Codex) arma tu bot de práctica y corre contra él herramientas abiertas de pruebas de seguridad. Solo se ataca lo que armaste vos, con datos ficticios.

Cómo se cursa: Cada semana tenés material grabado para estudiar y practicar antes de la clase. La clase es en vivo y sirve de apoyo: resolvemos dudas, explicamos lo que se complica y enseñamos lo justo. Las clases en vivo no se graban.

Soporte incluido: Todas las capacitaciones incluyen 3 meses de soporte, ayuda y acompañamiento ante dudas, desde la última clase.

Lo que te llevas

- Un bot de prueba al que le hiciste inyección directa e indirecta, y su versión rediseñada.
- Un mapa de qué puede hacer cada agente tuyo y cuáles permisos le sobran.
- Un webhook de WhatsApp con la firma validada y autenticación en tus flujos.
- Una suite de casos de abuso que corre sola cada vez que cambiás algo.
- [illegible] que un abuso no te [illegible] una factura.



Dicta Julián Bagilet
Más de 15 años en tecnología. Estudio de software en Villa General Belgrano, Córdoba.

Temario

1

SEMANA 1 SEGURIDAD

Cómo se rompe un bot con IA: el laboratorio

Armás un bot de práctica con datos ficticios de propiedades y lo atacás para ver, en vivo, dónde cede.

QUÉ VEMOS

- Los riesgos de la lista OWASP para aplicaciones con modelos de lenguaje 2025: inyección de instrucciones, fuga de información, exceso de agencia y consumo sin límite.
- Inyección directa: el usuario le pide al bot que ignore sus reglas. Por qué un modelo no distingue entre una instrucción y un dato, y por qué la lista de frases prohibidas no sirve.
- Inyección indirecta: un PDF, un mensaje o una ficha con texto oculto que el bot obedece. Lo armás con un RAG de práctica y medís el efecto.
- La fuga del system prompt: por qué nunca va una clave ni una regla de negocio ahí, y cómo rediseñar para que no haya secreto que filtrar.
- Los casos reales: Air Canada fue responsabilizada por lo que dijo su chatbot en 2024 (según una reseña secundaria). Lo que tu bot promete, te obliga.

Al terminar la semana: Terminás la semana con tu bot de práctica atacado, con el informe de dónde cedió y con la primera versión corregida.

2

SEMANA 2 SEGURIDAD

Permisos del agente y la regla de dos

Mapeás qué puede leer, escribir y enviar cada agente tuyo, y quitás lo que le sobra antes de que alguien lo use.

QUÉ VEMOS

- La "trifecta letal": acceso a datos privados, contenido no confiable y capacidad de comunicar hacia afuera. Con las tres juntas, un texto escondido alcanza para sacar datos.
- La regla de dos de Meta (octubre de 2025): un agente no debería reunir más de dos de esas tres cosas; si la tarea las exige, entra una persona a aprobar.
- Mínimo privilegio con casos de inmobiliaria: el asistente que resume mails no debería poder enviarlos, y el que atiende consultas no debería leer contratos.
- La autorización va en el sistema de atrás y no en el modelo: el bot pide y tu código decide. Aprobación atada a una acción y de un solo uso para lo irreversible.
- Un caso real: un ticket de soporte con instrucciones ocultas sacó una tabla de tokens de una base de Supabase conectada con la clave de servicio (julio de 2025).

Al terminar la semana: Terminás la semana con el mapa de permisos de cada uno de tus agentes y con tu versión corregida aplicando la regla de dos.

Temario (continuación)

3

SEMANA 3 SEGURIDAD

Webhooks, n8n, Make y Apps Script

Tus flujos reciben datos de afuera. Probás qué pasa sin firma ni autenticación, y los dejás cerrados.

QUÉ VEMOS

- Meta firma cada mensaje de WhatsApp con un encabezado de firma, pero validarlo es opcional. Quien no lo valida acepta datos de cualquiera: lo probás con un envío falso y luego lo cerrás.
- n8n: el nodo de webhook ofrece autenticación por encabezado, JWT o Basic, y una lista de IP. Con la lista vacía puede invocarlo cualquiera.
- Las vulnerabilidades de n8n de 2025 y 2026 como ejemplo de por qué se actualiza: una de ellas (CVE-2025-68613) entró al catálogo de explotadas de CISA en marzo de 2026.
- Apps Script: "ejecutar como yo" actúa con tu identidad aunque entre otro, y su token nunca debe viajar al navegador.
- Permisos de OAuth: en agosto de 2025 tokens robados de una integración tercera sirvieron para exportar datos de instancias de Salesforce. Pedir lo mínimo y rotar.

Al terminar la semana: Terminás la semana con tu webhook de WhatsApp con firma validada y con tus flujos con autenticación, versión revisada y lista de qué actualizar.

4

SEMANA 4 SEGURIDAD

Gasto, secretos, pruebas automáticas y registros

Ponés tope al gasto, sacás las claves del código y dejás pruebas de abuso que corren solas con cada cambio.

QUÉ VEMOS

- El abuso como costo: un atacante puede hacer que tu bot gaste. Límites de gasto mensual, alertas y topes por usuario y por conversación en Anthropic y OpenAI.
- Secretos: según GitGuardian, en 2025 se filtraron más de 28 millones de secretos nuevos en GitHub público. Variables de entorno, un escáner de secretos y rotación.
- Pruebas automáticas de abuso: garak, PyRIT y promptfoo son abiertas y se corren sobre tu bot de práctica y sobre nada más. Se repiten tras cada cambio.
- Qué registrar: cada llamada a una herramienta, cada pedido bloqueado y los intentos repetidos, con los datos personales tapados.
- Tu plan de incidente: si el bot filtra algo, quién lo apaga, a quién se avisa y cuándo.

Al terminar la semana: Terminás con tus topes de gasto activos, tus claves rotadas, tus pruebas de abuso corriendo y tu plan de incidente escrito.

Seguridad y alcance

Seguridad incluida en esta capacitación

- Pruebas solo sobre sistemas tuyos o con permiso escrito, con datos ficticios.
- La autorización la decide tu código, y toda acción irreversible la aprueba una persona.
- Topes de gasto, alertas y registros con los datos personales tapados.
- Nada de claves ni reglas de negocio dentro del prompt de un bot.
- Webhooks con firma validada y flujos con autenticación y actualizados.

Qué no se enseña ni se hace

- ✕ Atacar, probar o entrar a bots, cuentas o sistemas de otras personas o empresas: el artículo 153 bis del Código Penal sanciona el acceso sin autorización o excediéndola.
- ✕ Armar mensajes de engaño, clonar voces ni publicar paquetes con nombres falsos para probar un ataque.
- ✕ Asesoramiento legal: si dudás de un alcance, consultá a tu abogado.

Requisitos

Ayuda saber leer algo de código o haber hecho P6, pero no es excluyente. Una computadora y un plan pago de Claude o de Codex (OpenAI), no incluido en el precio. No se usan datos de clientes reales.

Inversión y formas de pago

Precio de la capacitación

$ 180.000

En vivo + material grabado · 15 cupos. Valores en pesos argentinos.

mercado pago

CUOTAS **3 cuotas de $ 60.000**
sin interés, con tarjeta, por Mercado Pago

TRANSFERENCIA **$ 153.000**
15% de descuento pagando por transferencia bancaria

Cómo inscribirte

1. **Escribinos**
Contanos qué querés lograr por WhatsApp.

2. **Confirmamos cupo y fecha**
Te avisamos apenas hay fecha y lugar en la cohorte.

3. **Reservá tu lugar**
Elegís cuotas o transferencia y recibís la confirmación por escrito.



WhatsApp +54 9 3546 56-6697

julianbagilet.com/capacitaciones/ · Instagram @mangopips_ok · Villa General Belgrano, Córdoba

Condiciones

Valores vigentes hasta el 31 de enero de 2027. Certificado de aprobación del estudio de Julián Bagilet, sin validez oficial. Derecho de arrepentimiento de 10 días corridos en compras a distancia por personas (Ley 24.240, art. 34). Si una cohorte no llega al mínimo de inscriptos, se reprograma o se devuelve el 100% de lo abonado. Los casos de las clases se explican como patrones, sin nombres ni datos de clientes.