

Agentic Evals

Objetivo del curso

Desarrollar evaluaciones sistemáticas para ingeniería de agentes, incluyendo RAG, agentes y voice agents, con el fin de medir calidad, detectar fallas, comparar modelos y tomar decisiones informadas para el despliegue de sistemas confiables en entornos reales.

Casos de estudio:

- RAG legal: Consulta de leyes/contratos; evalúa grounding, retrieval y hallucinations.
- Chatbot médico: Asistente de salud; evalúa exactitud, seguridad y riesgo en respuestas.
- RAG empresarial: Base de conocimiento interna; evalúa relevancia y cobertura de información.
- Voice AI (call center): Interacción voz a voz; evalúa STT, latencia y coherencia conversacional.
- Agente de gestión de pedidos (e-commerce): Automatiza pedidos con APIs; evalúa planificación, tools y manejo de errores.

Temas

1. Fundamentos de IA y Large Language Models (3 horas)

- 1.1. ¿Qué son los Large Language Models?
- 1.2. Cómo funcionan los LLMs (nivel general)
- 1.3. Tokens, embeddings, context window
- 1.4. Flujo Prompt → Modelo → Respuesta
- 1.5. Panorama de modelos: GPT, Claude, Llama, Gemini
- 1.6. Aplicaciones reales de LLMs
- 1.7. Software tradicional vs sistemas de IA

2. Python para evaluación y automatización de IA

- 2.1. Entornos virtuales y configuración
- 2.2. Consumo de APIs (requests)
- 2.3. Manejo de JSON y datasets de evaluación
- 2.4. Manejo de archivos y logs
- 2.5. Manejo de variables de entorno (.env)
- 2.6. Introducción a pytest
- 2.7. Estructura de tests reutilizables
- 2.8. Automatización de ejecuciones de evaluación

3. Fundamentos del Testing de IA

- 3.1. Testing tradicional vs testing de IA
- 3.2. Salidas deterministas vs probabilísticas



- 3.3. Retos únicos del testing en IA
- 3.4. Hallucinations (alucinaciones)
- 3.5. Sesgo y fairness
- 3.6. Pruebas de toxicidad
- 3.7. Sensibilidad al prompt
- 3.8. Riesgos de regresión
- 3.9. Model drift
- 3.10. Fundamentos de privacidad y cumplimiento
- 4. Evals para LLMs con DeepEval**
 - 4.1. Introducción a DeepEval
 - 4.2. Diseño de casos de prueba
 - 4.3. Evaluadores y métricas
 - 4.4. LLM-as-a-judge
 - 4.5. Evaluación basada en reglas
 - 4.6. Creación de golden datasets
 - 4.7. Flujos de validación automatizados
 - 4.8. Métricas: relevancia, faithfulness, correctness, hallucination, toxicity, bias, latency
- 5. Evals para RAG con RAGAS**
 - 5.1. ¿Qué es RAG?
 - 5.2. Flujo retriever + generator
 - 5.3. Validación de chunking
 - 5.4. Calidad de embeddings
 - 5.5. Validación de bases vectoriales
 - 5.6. Pruebas de groundedness
 - 5.7. Evaluación del retrieval
 - 5.8. Métricas: faithfulness, context precision, context recall, answer relevancy
- 6. Evals para Agentes de IA**
 - 6.1. Introducción a sistemas agentic
 - 6.2. Arquitectura: planner + executor
 - 6.3. Validación de uso de herramientas (tools)
 - 6.4. Workflows multi-paso
 - 6.5. Manejo de memoria y contexto
 - 6.6. Evaluación de razonamiento multi-hop
 - 6.7. Integración con RAG (agentic RAG)
 - 6.8. Validación de selección de herramientas
 - 6.9. Evaluación de consistencia de respuestas
 - 6.10. Pruebas de fallo, fallback y recuperación
 - 6.11. Métricas: task completion, tool correctness, argument correctness, turn relevancy, conversation completeness
- 7. Evals para Prompts y Regresión con Promptfoo**
 - 7.1. Introducción a Promptfoo
 - 7.2. Aserciones basadas en YAML
 - 7.3. Validación de outputs
 - 7.4. Comparación entre modelos
 - 7.5. Testing de regresión
 - 7.6. Verificación de prompts críticos y safety

8. Evals para Voice AI

- 8.1. Arquitectura de sistemas de voz (STT → LLM → TTS)
- 8.2. Pruebas de speech-to-text
- 8.3. Validación de respuestas del LLM
- 8.4. Pruebas de text-to-speech
- 8.5. Evaluación de intent y comprensión
- 8.6. Medición de latencia
- 8.7. Pruebas de interrupción, fallback y flujo conversacional

Herramientas a aprender a usar:

- Python- Base para automatización y evaluación de IA.
- DeepEval - Evaluación de LLMs y agentes.
- RAGAS - Evaluación de sistemas RAG.
- Promptfoo - Testing de prompts y regresión.
- pytest - Framework de testing automatizado.
- Requests - Consumo de APIs.
- python-dotenv - Manejo de variables de entorno.

¿Para quién es este curso?

Este curso está diseñado para perfiles técnicos que quieren especializarse en la evaluación sistemática de sistemas de IA:

- **QA Engineers, Software Testers y SDETs, Profesionales de testing y automatización** que quieren llevar su experiencia al siguiente nivel: evaluar LLMs, RAG y agentes con métricas reales (faithfulness, hallucinations, tool correctness) usando DeepEval, RAGAS, Promptfoo y pytest.
- **Desarrolladores e Ingenieros de IA/ML.** Quienes construyen aplicaciones con LLMs, RAG o agentes, y necesitan validarlas de forma rigurosa —no solo "que funcione"— para detectar fallas, comparar modelos y desplegar con confianza en producción.
- **Profesionales en transición a roles de evaluación de IA.** Personas con bases en programación que buscan moverse a posiciones como AI Tester o LLM Evaluation Engineer, y que necesitan un método estructurado y profesional para evaluar sistemas agenticos.
- **Entusiastas de la tecnología con experiencia básica en IA.** Que ya han experimentado con herramientas o modelos de IA y buscan aprender a evaluarlos de forma estructurada y profesional.

Requisitos

- Una computadora de escritorio o laptop de 64 bits (6GB de RAM).
- Opcional: Si desea el alumno replicar el contenido en su local se requiere:
 - o Una laptop con mínimo 8GB RAM y con GPU Nvidia mínimo GTX 1060 Ti de 8GB de memoria, cualquier RTX con 8Gb de memoria.
- Conexión a Internet con un ancho de banda mayor o igual a 5 Megas.
- Para el curso puedes usar Windows, Mac o Linux como tu gustes.
- Asegúrate de darnos tus datos de contacto al comprar (tu nombre completo, tu e-mail y tu número telefónico).
- (opcional, pero recomendable): Tener cuenta en [Colab](#) (gmail), [Kaggle](#) y [HuggingFace](#).

Conocimientos Previos

- Conocimientos básicos de matemáticas (nivel preparatoria o primeros semestres de una carrera STEM). No es necesario un enfoque teórico profundo; el curso es práctico.
- Conocimientos básicos de programación (cualquier lenguaje). Capacidad de leer y entender código sencillo.
- Interés en inteligencia artificial y sistemas basados en LLMs. Se recomienda contar con una noción básica o haber tenido un primer acercamiento a herramientas de IA.

Reconocimiento

Al finalizar el curso, recibe un reconocimiento con valor curricular:



Registro ante la Secretaría del Trabajo y Previsión Social: **ZAGE-810930-FW2-0005**

Lo que querías saber:

Las clases son en línea vía Zoom, cada clase se va grabando y te damos acceso a las grabaciones por 5 meses.

Reconocimiento digital ante la Secretaría del Trabajo (STPS-México)

Reserva con solo \$800 MXN y paga el resto en la 4ta semana

Precio total \$2500 MXN

Paga mediante Paypal/Stripe/Transferencia Bancaria. Contacta a Claudia Montañaño wa.me/5215539940156.

Inicio del curso y fin del curso

Lunes 20 de Julio 2026

Lunes 21 de Septiembre 2026 (No hay clases el 16 de septiembre)

Horarios

36 horas, 4 horas la semana (Horario de la Ciudad de México)

- Lunes y Miércoles a las 20 hrs

Instructor

Dr. Uriel Escalona, ver [aquí](#) sus reseñas.

Testimonios

Hemos capacitado a más de 2000 personas en IA. [Aquí encontrarás lo que dicen.](#)
[Ve hasta abajo.](#)

Medios de Pago

Paga con tarjetas de crédito/débito (Mastercard, Visa, Amex) en la página

www.actumlogos.com

Transferencia/deposito bancario -> [Datos bancarios](#)

Solicita tu pago por QR+CoDi/Paypal -> wa.me/5215539940156 Claudia Montañaño, Mail: hola@actumlogos.com