CTIC UNI

UNIVERSIDAD NACIONAL DE INGENIERÍA

OFICINA DE TECNOLOGÍAS DE LA INFORMACIÓN

Programa de Iniciación Tecnológica PIT - 2026

CERTIFICADO

Otorgado a

MANOLO CANALES CUBA

Por haber aprobado el curso de

GRANDES MODELOS DE LENGUAJE (LLMS): CONSTRUYENDO UN AGENTE INTELIGENTE

Realizado del 20 de enero al 31 de enero del presente, con una duración de 16 horas.

Lima, febrero de 2026

Mag. Ing. Rubén Arturo Borja Rosales
JEFE OTI - UNI

N° Certificado: 018 - 0002978

FIRMA DIGITAL

Firmado digitalmente por: BORJA ROSALES Ruben Arturo FAU 20169004359 hard
Motivo: En señal de conformidad
Fecha: 25/02/2026 15:42:02-0500

UNIVERSIDAD NACIONAL DE INGENIERÍA

OFICINA DE TECNOLOGÍAS DE LA INFORMACIÓN

CTIC UNI

Temario del Curso:

GRANDES MODELOS DE LENGUAJE (LLMS): CONSTRUYENDO UN AGENTE INTELIGENTE

- Fundamentos de los LLMS.
- Recordar cómo la estadística del lenguaje evoluciona a Transformers.
- Comprender Tokenización, Embeddings y mecanismo de atención.
- Analizar la complejidad computacional y los cuellos de botella de un Transformer.
- Ecosistema OPEN-SOURCE Y serving local con OLLAMA.
- Ajuste eficiente: LORA y QLORA.
- Entender la idea de descomposición de rango bajo.
- Calcular reducción de parámetros y VRAM.
- Afinar un Llama-3-8B en dominio legal con PEFT + bitsandbytes.
- Validar con métricas de perplexity y exact match.
- RAG Y HYBRID-RAG.
- Revisar por qué RAG mitiga alucinaciones.
- Crear embeddings con sentence-transformers, almacenar en Chroma o Qdrant.
- Implementar búsqueda híbrida (similaridad + Bm25).
- Combinar resultados con prompt templates y “stuff merging”.
- Diseño de agentes y orquestación con N8N / LANGCHAIN.
- Arquitectura agentic: plan-act-react.
- Definir Tools, Memory y manejo de contexto.
- Crear un flujo en n8n que:
- Reciba una consulta.
- Llame al RAG de la sesión. Envíe respuesta a un canal Slack o Telegram.
- Despliegue ON-PREMISE, Evaluación Y GUARDRAILS.
- Empaquetar el agente en Docker (Ollama + n8n + vector DB).
- Configurar autenticación, límites de tasa y auditoría de logs.
- Introducir guardrails (Prompt checking, valid JSON schemas).
- Métricas de producción: Latency, token usage, satisfaction score.
- Ética: Privacidad, sesgos y control de datos.

Nota final: 18 (Dieciocho)

N° Certificado: 018 - 0002978