

Diplomado en ARTIFICIAL INTELLIGENCE AND LARGE LANGUAGE MODELS FOR FINANCIAL MARKETS (Versión en línea)

ITAM

Extensión Universitaria
y Desarrollo Ejecutivo



Nota: Este programa es en la modalidad en línea a través de la herramienta Zoom. Se requiere que el participante cuente con computadora, laptop, tablet, teléfono inteligente o cualquier otro dispositivo que permita reproducir audio y video y una buena conexión a internet.

Las clases serán en tiempo real en los días y horario publicados. Las sesiones no serán grabadas y el participante sólo tendrá acceso a las sesiones del curso en el grupo al cual se haya inscrito.



Coordinador:
M.F. José Jorge Ramírez Olvera

Con una amplia y reconocida trayectoria en el medio Financiero y Académico de más de 15 años, José Jorge Ramírez Olvera es coordinador de los Diplomados en Derivados Financieros, Productos Financieros Estructurados, Data Science and Machine Learning Applied to Financial Markets, Modelos Cuantitativos Aplicados a Finanzas y Trading y Artificial Intelligence and Large Language Models for Financial Markets, impartidos en el Instituto Tecnológico Autónomo de México (ITAM) y profesor en otros programas de Maestría y Licenciatura en las principales universidades del país. En el sector financiero, la experiencia de José Jorge incluye el control de riesgos financieros hasta la estructuración de nuevos productos financieros basados en innovación, tendencias y oportunidades de mercado. Dirige el área de Productos Estructurados en Analysis, socio consultor en Fundamenta Structuring, QAID y Analysis Nabla, también se ha encargado de las mesas de estructuración en distintas instituciones financieras, entre las que destacan: ING Bank, Bank of America – Merrill Lynch y Cantor Structuring (filial de Cantor Fitzgerald). Su especialidad se centra en la cobertura de derivados exóticos, desarrollo de metodologías de valuación de instrumentos financieros y programas de emisión internacionales y locales. Principal contacto con clientes (institucionales y corporativos), intermediarios financieros y autoridades. José Jorge es licenciado en Actuaría, licenciado en Matemáticas Aplicadas y maestro en Finanzas por el ITAM.

Objetivo General

Formar especialistas capaces de **diseñar, implementar y evaluar soluciones avanzadas de inteligencia artificial y Large Language Models (LLMs) aplicadas al sector financiero**, mediante el uso intensivo de Python y herramientas modernas de ciencia de datos, aprendizaje de máquina e inteligencia artificial generativa.

A lo largo del diplomado, los participantes desarrollarán soluciones basadas en **modelos de lenguaje de gran escala**, incorporando técnicas contemporáneas como **preentrenamiento, fine-tuning, Retrieval-Augmented Generation (RAG)**, así como estrategias de despliegue, evaluación y control de alineamiento de modelos, con especial énfasis en su uso responsable y efectivo en entornos financieros reales. El programa está diseñado con un **nivel avanzado**, asumiendo que los participantes cuentan con experiencia previa en programación, análisis

cuantitativo y finanzas. El desarrollo del diplomado combina fundamentos teóricos sólidos con una orientación marcadamente práctica, en la que los asistentes trabajarán con **casos reales del sector financiero**, integrando datos estructurados y no estructurados para resolver problemas complejos relacionados con **análisis financiero, gestión de riesgos, cumplimiento normativo, auditoría, operaciones financieras y automatización documental**.

Al finalizar el diplomado, los participantes no solo comprenderán el funcionamiento interno y las limitaciones de los LLMs, sino que serán capaces de **integrarlos de manera crítica y profesional en procesos financieros**, evaluando su impacto, riesgos y beneficios desde una perspectiva técnica, cuantitativa y regulatoria, acorde con las exigencias actuales de la industria financiera. Además, será capaz de:

- Diseñar arquitecturas basadas en LLMs, modelos generativos y sistemas híbridos (ML + LLMs) para aplicaciones financieras.
- Integrar datos estructurados y no estructurados (mercados, reportes, regulación, noticias, contratos).
- Construir agentes inteligentes para análisis financiero, riesgo, trading, compliance y soporte a decisiones.
- Evaluar riesgos, sesgos, limitaciones y aspectos regulatorios del uso de IA en finanzas.
- Traducir problemas financieros complejos en soluciones computacionales robustas y escalables.

Dirigido a:

A profesionistas del sector financiero del país (bancos, casas de bolsa, sociedades de inversión, afores, casas de cambio, operadores de derivados, socios liquidadores, arrendadoras financieras, empresas de factoraje financiero, compañías de seguros y fianzas, entidades gubernamentales, entidades reguladoras, consultorías, etc.) además de:

- Científicos de datos y *quantitative analysts* que trabajan con información financiera.
- Desarrolladores y arquitectos de soluciones en fintech y regtech.
- Consultores financieros y tecnológicos.

El diplomado está diseñado para participantes con una **formación técnica y cuantitativa sólida**, por lo que se asume que los aspirantes cuentan con experiencia previa en programación y finanzas. En particular, se espera que los participantes posean:

- **Conocimientos sólidos de programación en Python**, incluyendo el uso habitual de librerías para análisis de datos y aprendizaje de máquina como Pandas, NumPy y scikit-learn, así como experiencia con al menos un framework de deep learning (por ejemplo, TensorFlow o PyTorch).
- **Bases firmes en matemáticas financieras**, así como en **cálculo, probabilidad y estadística**, necesarias para la correcta comprensión y evaluación de modelos avanzados.
- **Conocimiento de productos financieros básicos**, tales como bonos, derivados plain vanilla y nociones de construcción y análisis de portafolios.
- **Disponibilidad de equipo de cómputo adecuado**, ya sea una computadora portátil con GPU dedicada, o bien acceso a servicios de cómputo en la nube con soporte de GPU, para el entrenamiento y experimentación con modelos de mayor complejidad.
- De manera ideal, haber cursado el **Diplomado en Data Science and Machine Learning Applied to Financial Markets**, o bien contar con **experiencia profesional o académica equivalente** en ciencia de datos aplicada a finanzas.

Estos requisitos permiten asegurar que el desarrollo del diplomado se mantenga en un nivel avanzado, con un enfoque en aplicaciones reales y profundización técnica, sin necesidad de cubrir contenidos introductorios.

Beneficios:

- Aplicar LLMs y modelos generativos a problemas reales del sector financiero (riesgo, compliance, auditoría, operaciones y análisis).
- Dominar técnicas actuales como fine-tuning y RAG para crear soluciones robustas con datos estructurados y no estructurados.
- Construir agentes inteligentes para automatizar investigación, soporte a decisiones y flujos financieros críticos.
- Elevar tu perfil cuantitativo-tecnológico con un enfoque intensivo en Python, ML/DL y herramientas modernas de IA aplicada.
- Implementar con estándares institucionales, priorizando evaluación, trazabilidad, gobierno de modelos y uso responsable en entornos regulados.

Módulo 1: Advanced Python, Machine Learning Stack and LLM Tooling Foundations

Objetivo:

Homogeneizar y consolidar las bases técnicas avanzadas en Python, machine learning y deep learning necesarias para el desarrollo profesional de aplicaciones basadas en LLMs y sistemas inteligentes en el sector financiero.

Temario:

1. Flujo profesional de trabajo en Python
 - a. Python avanzado aplicado a proyectos de IA
 - b. Jupyter Notebooks, Google Colab y entornos virtuales
 - c. Reproducibilidad, experimentación y documentación
 - d. Estructura de proyectos profesional (src/, config, logging)
 - e. Gestión moderna de dependencias y empaquetado (pyproject.toml, Poetry/uv, lockfiles)
 - f. Calidad de código: ruff/black, pre-commit, type hints (mypy opcional)
 - g. Testing con pytest, coverage y mocks
 - h. Manejo seguro de secretos y configuración (dotenv + buenas prácticas)
2. Programación orientada a objetos y diseño de software
 - a. Clases, herencia y composición
 - b. Clean code y patrones de diseño
 - c. Estructuración de proyectos financieros complejos
 - d. Principios de diseño para pipelines de datos/ML (interfaces, modularidad, I/O contracts)
3. Stack científico y ML clásico en Python
 - a. NumPy y vectorización
 - b. Pandas para análisis financiero
 - c. Visualización con Matplotlib
 - d. Scikit-learn: pipelines, validación cruzada y métricas
 - e. Validación de datos (pandera) y serialización eficiente (Parquet)
 - f. Tracking de experimentos (MLflow)
4. Introducción profunda a PyTorch
 - a. Tensores, autograd y nn.Module
 - b. Redes neuronales simples y MLPs

- c. Entrenamiento y evaluación
- 5. Ecosistema Hugging Face y control de versiones
 - a. Transformers, Datasets y Hub
 - b. Git y GitHub para versionado de código, modelos y aplicaciones
 - c. Buenas prácticas de versionado de datos/modelos (DVC)
- 6. Fundamentos de contenedores para reproducibilidad
 - a. Docker (build/run, Dockerfile, .dockerignore, reproducibilidad)
 - b. Ejecución local reproducible de inferencia (mini-API o script)
- 7. Casos financieros introductorios
 - a. Clasificación de noticias financieras
 - b. Análisis preliminar de riesgo y sentimiento

Módulo 2: Historical Context, Transformers and Core LLM Architecture

Objetivo:

Comprender la evolución histórica del procesamiento de lenguaje natural hasta los LLMs modernos y dominar en profundidad la arquitectura Transformer como base de las aplicaciones generativas en finanzas.

Temario:

- 1. Evolución del NLP
 - a. Sistemas basados en reglas
 - b. Modelos estadísticos
 - c. Modelos neuronales
- 2. Primeros modelos de lenguaje
 - a. N-grams
 - b. Word embeddings: Word2Vec, GloVe
- 3. Surgimiento de los LLMs
 - a. Preentrenamiento masivo
 - b. Cambios de escala y capacidades emergentes
- 4. Arquitectura Transformer
 - a. Attention mechanism
 - b. Multi-head attention
 - c. Positional encoding
 - d. Feed-forward layers
 - e. Tokenización moderna (BPE), batching/packing y efectos en contexto
 - f. Complejidad, memoria y nociones prácticas de KV cache
- 5. Variantes de Transformers
 - a. GPT, BERT, T5
 - b. Modelos encoder, decoder y encoder–decoder
 - c. Modelos de long-context y estrategias prácticas: segmentación/sliding window
- 6. Implementación práctica en PyTorch
 - a. Construcción de un mini-Transformer
 - b. Aplicaciones financieras:

- i. clasificación de noticias,
- ii. detección de riesgo,
- iii. etiquetado de texto financiero
- iv. Introducción a embeddings para búsqueda semántica

Módulo 3: Pretraining, Fine-Tuning, Prompt Engineering and RAG in Finance

Objetivo:

Desarrollar la capacidad de adaptar y especializar LLMs a tareas financieras mediante técnicas modernas de fine-tuning, prompt engineering, alineamiento y Retrieval-Augmented Generation.

Temario:

1. Objetivos de preentrenamiento
 - a. Masked Language Modeling
 - b. Next Sentence Prediction
2. Transfer learning y adaptación de dominio financiero
3. Estrategias de fine-tuning
 - a. Full fine-tuning
 - b. Métodos eficientes: LoRA, PEFT
 - c. Cuando no hacer fine-tuning: comparativo costo/beneficio
4. Construcción de datasets financieros
 - a. Documentos regulatorios
 - b. Noticias, reportes, transacciones
 - c. Gobernanza de datos: versionado de corpus, licencias, PII/redacción y retención
5. Prompt engineering
 - a. Few-shot y one-shot prompting
 - b. Chain-of-thought
 - c. Prompt hacking e inyección
 - d. Salidas estructuradas (JSON), validación y guardrails de formato
 - e. Diseño de prompts con controles (plantillas, roles, políticas internas)
6. Instruction tuning y alineamiento
 - a. SFT
 - b. RLHF y DPO (visión conceptual)
7. Retrieval-Augmented Generation (RAG)
 - a. Embeddings y vector databases
 - b. Chunking y metadata
 - c. Búsqueda híbrida y re-ranking
 - d. GraphRAG y recuperación avanzada basada en grafos
 - e. Normativa financiera (CNBV, Banxico, etc)
 - f. Evaluación de RAG: retrieval metrics (recall@k), groundedness/faithfulness y análisis de errores
 - g. Estrategias de citación y trazabilidad de fuentes en respuestas
8. Laboratorios financieros
 - a. Asistentes de research
 - b. Q&A sobre reportes financieros
 - c. Compliance y documentación regulatoria
 - d. Pruebas adversarias: prompt injection, data exfiltration y mitigaciones

Módulo 4: Agentic AI, Reinforcement Learning and Intelligent Financial Decision Systems

Objetivo:

Diseñar sistemas de IA agente capaces de planear, razonar y ejecutar tareas complejas dentro de flujos financieros, manteniendo control, trazabilidad y cumplimiento regulatorio.

Temario:

1. De GenAI a IA agente
 - a. Definición y motivación
 - b. Diferencias con chatbots tradicionales
2. Componentes de un agente
 - a. Memoria
 - b. Planeación
 - c. Uso de herramientas
 - d. Políticas de memoria: qué guardar / qué no guardar (compliance + privacidad)
3. Frameworks de agentes en Python
 - a. LangChain
 - b. LlamaIndex
4. Arquitecturas de agentes
 - a. Agentes simples
 - b. Arquitecturas multi-agente
 - c. Patrones de control: approval gates, límites de herramientas, sandboxing
 - d. Observabilidad del agente: trazas, métricas, costos por tool-call
5. Diseño de herramientas financieras
 - a. APIs de mercado
 - b. Bases de datos SQL
 - c. Sistemas internos
 - d. Conectores a documentos regulados y RAG como “tool” (retrieval tool)
6. Agentes por función financiera
 - a. Front office
 - b. Middle office / Riesgos
 - c. Back office / operaciones
7. Introducción a Reinforcement Learning
 - a. RL en finanzas (visión ejecutiva)
 - b. Riesgos del RL en trading
8. Control, seguridad y compliance
 - a. Human-in-the-loop
 - b. Logging, auditoría y explicabilidad
 - c. Mitigación de tool-injection y verificación de acciones (policy checks)

Módulo 5: GPU-Accelerated Computing and Large-Scale Financial AI Infrastructure

Objetivo:

Capacitar al participante para diseñar y ejecutar soluciones financieras intensivas en cómputo utilizando GPUs, comprendiendo arquitectura, costos e impacto institucional.

Temario:

1. Contexto estratégico del cómputo acelerado
 - a. Fin de la Ley de Moore
 - b. Necesidades computacionales en finanzas
2. Arquitectura de GPU
 - a. CUDA cores
 - b. Memoria y latencias
 - c. Throughput
3. Ecosistema NVIDIA
 - a. CUDA
 - b. cuBLAS, cuRAND, cuDNN
4. Python en GPU
 - a. Numba
 - b. CuPy
 - c. RAPIDS: cuDF y cuML
5. Integración con frameworks de DL
 - a. PyTorch
 - b. Eficiencia práctica para LLMs: checkpointing, batch packing, mixed precision
 - c. Inferencia eficiente: cuantización (visión práctica), throughput/latencia, KV cache (aplicado)
 - d. Introducción a entrenamiento/inferencia distribuida (conceptual + demo)
6. Patrones de paralelismo financiero
 - a. Monte Carlo masivo
 - b. VaR y Expected Shortfall intradía
 - c. Optimización de portafolios de gran dimensión
7. Casos de estudio
 - a. xVA en GPU
 - b. Backtesting acelerado
 - c. High-Frequency Trading (HFT)
8. Infraestructura y costos
 - a. On-premise vs nube
 - b. Diseño de AI risk engines
 - c. Costeo de GPU en nube: estimación por experimento (benchmark + costo aproximado)

Módulo 6: LLMOps, Governance, and Deployment

Objetivo:

Integrar los conocimientos del diplomado en una arquitectura institucional robusta, considerando despliegue, monitoreo, gobierno de modelos y culminando con un proyecto final aplicado.

Temario:

1. Arquitecturas tipo “AI Factory”
 - a. Datos → cómputo → modelos → agentes → consumo
2. MLOps y LLMOps
 - a. Versionado de datos y modelos
 - b. Monitoreo de desempeño y drift
 - c. Evaluación continua: calidad, groundedness, retrieval health, regresiones

- d. Dataset/model cards + data lineage como entregables
- 3. Optimización de inferencia
 - a. Cuantización
 - b. Distillation
 - c. Pruning
- 4. Model serving
 - a. APIs
 - b. Contenedores
 - c. Costos y escalabilidad
 - d. Docker avanzado: multi-stage builds, imágenes slim, runtime vs build
 - e. CI/CD: build/test/scan/deploy (GitHub Actions)
 - f. Observabilidad: logs estructurados, métricas, tracing (OpenTelemetry visión práctica)
 - g. Rate limiting, cuotas y control de costos por usuario/proceso
 - h. APIs, FastAPI, Gradio y Streamlit
- 5. Cloud & Production GenAI
 - a. Nociones de despliegue en nube (compute, storage, networking)
 - b. Gestión de secretos (KMS/Secrets Manager o equivalente)
 - c. IAM y mínimos privilegios para servicios de IA
 - d. Vector DB en producción: backups, reindex, rollbacks, versionado de embeddings
- 6. Seguridad, alineamiento y ética
 - a. Bias y fairness
 - b. Riesgos regulatorios
 - c. Riesgos técnicos explícitos de LLMs (hallucinations, prompt injection, data leakage)
- 7. Gobierno de modelos
 - a. Model Risk Management
 - b. Documentación para auditoría
 - c. Playbooks de incidentes (prompt injection, data leakage, degradación RAG)
 - d. Controles y evidencias para auditoría (logs, trazas, approvals)
- 8. Multimodal LLMs (texto–imagen–audio)