

## Ficha técnica

# Programación de modelos de lenguaje de gran tamaño con Azure Open AI.

Programación conversacional e ingeniería de prompt con LLM.

College

### Autores

Francesco Esposito

### Revisión técnica

Dr. Andrés Mauricio Castillo Robles

*Australian National Phenome Centre*

Dr. Jorge Adrián Lázaro Hernández

*Customs City Global Solutions Inc.*



## Objetivo



**Programación de modelos de lenguaje de gran tamaño con Azure Open AI** explora la aplicación práctica de los LLMs existentes en el desarrollo de aplicaciones versátiles de dominio empresarial. En esencia, un LLM es un modelo de ML entrenado en extensos datos de texto, lo que le permite comprender y generar lenguaje humano.

Este libro se dirige a una audiencia diversa, incluyendo a programadores, entusiastas de ML, profesionales de la computación en la nube, y aquellos interesados en el procesamiento del lenguaje natural, con un énfasis específico en el aprovechamiento de los servicios de Azure para programar LLMs.

ISBN-EPUB: 978-607-32-6007-7

ISBN-ETEXT: 978-607-32-6047-3

Formato: epub

Edición: 1ª edición

Nivel de enseñanza: College

Área/Subárea: Computación

Fecha de publicación: Agosto

Copyright: 2024



## Características (promotips)

Para comprender plenamente el valor de un libro de programación sobre LLMs, se recomienda una combinación de experiencia en programación, conciencia de ML y comprensión lingüística. Algunos de los requisitos previos son: competencia en conceptos fundamentales de programación y familiaridad con los fundamentos de ML, un conocimiento práctico de los lenguajes de programación y *frameworks* como *Python* y posiblemente *ASP.NET Core*, serán útiles, al igual que una apreciación por la importancia del procesamiento clásico del lenguaje natural en el contexto de los dominios empresariales.

Para abordar los conocimientos sobre estos modelos, este libro se centra en tres aspectos clave:

- Los primeros tres capítulos profundizan en los escenarios para los cuales un LLM es efectivo e introducen herramientas esenciales para elaborar soluciones sofisticadas. Estos capítulos proporcionan ideas sobre la programación conversacional y la incitación como un nuevo enfoque de codificación avanzado, pero estructurado.
- Los dos capítulos siguientes enfatizan patrones, marcos y técnicas para desbloquear el potencial de la programación conversacional. Esto implica el uso de lenguaje natural en el código para definir flujos de trabajo, con la aplicación basada en LLM orquestando las API existentes.
- Los tres capítulos finales presentan ejemplos de demostración concretos, de principio a fin, con *Python* y *ASP.NET Core*. Estas demostraciones muestran interacciones progresivamente avanzadas entre la lógica, los datos y los procesos empresariales existentes:

- » En la primera demostración, aprenderá cómo tomar texto de un correo electrónico y elaborar un borrador adecuado para una respuesta.
- » En la segunda demostración, aplicará un patrón de generación aumentada de recuperación (RAG) para formular respuestas a preguntas basadas en el contenido del documento.
- » En la tercera demostración, aprenderá cómo construir una aplicación de reserva de hotel con un chatbot que utiliza una interfaz conversacional para determinar las necesidades del usuario (fechas, preferencias de habitación, presupuesto) y realizar reservas de manera fluida, según el estado del sistema subyacente, sin utilizar elementos de interfaz de usuario fijos o controles de entrada de datos formateados.

Utilice LLM para crear mejores aplicaciones de software empresarial, comunicándose de forma autónoma con los usuarios y optimizando las tareas comerciales con aplicaciones para hacer que la interacción entre humanos y computadoras sea fluida y natural.

### Material de apoyo para docentes

Notebooks de *Python* y *Políglota* que contienen el código presentado en la parte inicial del libro.

Para mayor información, consulte a su Asesor de Servicios de Aprendizaje de Pearson o escriba a [hedsoporte@pearson.com](mailto:hedsoporte@pearson.com)



## Ventajas competitivas

La tecnología detrás de aplicaciones asombrosas como ChatGPT se llama modelo de lenguaje de gran tamaño (LLM). Los LLM añaden una capacidad crucial a la IA: la capacidad de generar contenido además de clasificar y predecir, por lo que representan un cambio de paradigma, elevando el nivel de comunicación entre humanos y computadoras.

El experto en Inteligencia Artificial Francesco Esposito le ayudará a través de la obra a:

- ★ Comprender la historia de los grandes modelos de lenguaje y la programación conversacional.
- ★ Aplicar indicaciones como una nueva forma de codificación
- ★ Aprender técnicas de indicaciones básicas y casos de uso fundamentales
- ★ Diseñar indicaciones avanzadas, incluida la conexión de LLM a datos y llamadas de funciones para crear motores de razonamiento.
- ★ Utilizar el lenguaje natural en el código para definir flujos de trabajo y orquestar las API existentes
- ★ Dominar los marcos externos de LLM
- ★ Evaluar las preocupaciones de seguridad, privacidad y precisión de la IA responsable

- ★ Explorar el panorama regulatorio de la IA
- ★ Construir e implementar un asistente personal
- ★ Aplicar un patrón de recuperación de generación aumentada (RAG) para formular respuestas basadas en una base de conocimientos.
- ★ Construir una interfaz de usuario conversacional

La obra está dirigida a los arquitectos de software, los desarrolladores líderes y las personas con experiencia en programación, especialmente aquellos familiarizados con lenguajes como Python y posiblemente C# (para ASP.NET Core), encontrarán el contenido de este libro accesible y valioso. También es útil para los profesionales de la nube y de TI con interés en el uso de servicios en la nube (específicamente Microsoft Azure) o en aplicaciones sofisticadas y reales de lenguaje humano en **software**. Aún y cuando el contenido se centra principalmente en los servicios disponibles en la plataforma Microsoft Azure, los conceptos cubiertos son fácilmente aplicables a plataformas análogas.

Les invitamos a considerar que el libro no puede ser la guía definitiva para la IA y los LLMs, debido a la impresionante velocidad en la que evolucionan, además, por diseño, cada libro es un acto de aproximación, como una fotografía del conocimiento tomada en un momento específico en el tiempo.



## Introducción

### CAPÍTULO 1. El origen y un análisis de los grandes modelos de lenguaje

#### LLM de un vistazo

- Historia de los LLM
- Fundamentos de funcionamiento
- Casos de uso empresarial

#### Hechos sobre la programación conversacional

- El poder emergente del lenguaje natural
- Topología de los LLM
- Perspectiva a futuro

#### Resumen

### CAPÍTULO 2. Técnicas básicas de aprendizaje de prompts

#### ¿Qué es la ingeniería de prompts?

- Un primer vistazo a los prompts
- Distintas formas de alterar la respuesta
- Preparación para la ejecución del código

#### Técnicas básicas

- Escenarios de cero disparos
- Escenarios de pocos disparos
- Escenarios de cadena de pensamiento

#### Casos de uso fundamentales

- Chatbots
- Traducción

#### Limitaciones de los LLM

#### Resumen

### CAPÍTULO 3. Aprendizaje avanzado de ingeniería de prompts

#### ¿Qué hay más allá de la ingeniería de prompts?

- Combinar piezas
- Afinación

#### Llamada a función

- Estilo casero
- Estilo OpenAI

#### Hablar con los datos (separados)

- Cómo conectar datos a LLM
- Embeddings
- Almacenamiento de vectores
- Generación aumentada de recuperación

#### Resumen

### CAPÍTULO 4. Dominio de los marcos de lenguaje

#### La necesidad de un orquestador

- Conceptos de marco cruzado
- Puntos a considerar

#### LangChain

- Modelos, plantillas de prompts y cadenas
- Agentes

#### Conexión de datos

#### Semantic Kernel de Microsoft

- Plug-ins
- Datos y planificadores

#### Guía de Microsoft

- Configuración
- Características principales

#### Resumen

### CAPÍTULO 5. Preocupaciones de seguridad, privacidad y precisión

#### Panorama general

- IA responsable
- Equipo rojo
- Abuso y filtrado de contenido
- Alucinación y rendimientos
- Sesgo y equidad

#### Seguridad y privacidad

- Seguridad
- Privacidad

#### Evaluación y filtrado de contenido

- Evaluación
- Filtrado de contenido

#### Resumen

### CAPÍTULO 6. Construcción de un asistente personal

#### Panorama general de la aplicación web del chatbot

- Alcance
- Pila de tecnología

#### El proyecto

- Configuración del LLM
- Configuración del proyecto
- Integración del LLM
- Posibles extensiones

#### Resumen

### CAPÍTULO 7. Chatee con sus datos

#### Panorama general

- Alcance
- Pila de tecnología

#### ¿Qué es Streamlit?

- Una breve introducción a Streamlit
- Principales características de la interfaz de usuario
- Ventajas y desventajas en producción

#### El proyecto

- Configuración del proyecto y la interfaz de usuario base
- Preparación de datos
- Integración del LLM

#### Cómo ir más allá de este punto

- Posibles extensiones
- Generación aumentada de recuperación versus afinación

Posibles extensiones  
**Resumen**

## **CAPÍTULO 8. Interfaz de usuario conversacional**

---

**Panorama general**  
Alcance  
Pila de tecnología

**El proyecto**  
Configuración de una API mínima  
OpenAPI  
Integración del LLM  
Posibles extensiones

**Resumen**  
**Apéndice**