

MODELOS PERSONALIZADOS DE INTELIGENCIA ARTIFICIAL GENERATIVA: DESARROLLA TU PROPIO CHATBOT

SKU: N / A | Categorías: [ADMINISTRACIÓN Y GESTIÓN](#), [INFORMÁTICA Y COMUNICACIONES](#)

INFORMACIÓN DEL CURSO

Horas 20

Cod. 18

CURSO DE ESPECIALIZACIÓN

MODELOS PERSONALIZADOS DE INTELIGENCIA ARTIFICIAL GENERATIVA: DESARROLLA TU PROPIO CHATBOT

Objetivos

En un mundo donde la tecnología evoluciona a un ritmo acelerado, el dominio de los modelos de lenguaje de gran tamaño (LLMs) y las tecnologías RAG se están convirtiendo en una herramienta crucial para innovar y transformar industrias. Este curso está diseñado para proporcionar una comprensión profunda de los LLMs, explorando cómo estos modelos no solo comprenden y generan texto con una precisión impresionante, sino cómo se pueden crear soluciones específicas que van desde la mejora de sistemas de recomendación hasta la automatización de respuestas en servicios de atención al cliente.

- Identificar y explicar los componentes clave y la funcionalidad de los Modelos de Lenguaje de Gran Escala (LLMs) y la Recuperación de Información Generativa (RAG).
- Evaluar ejemplos de aplicaciones RAG y describir cómo se integran con LLMs para mejorar la recuperación de información.
- Aplicar técnicas de procesamiento para transformar datos no estructurados en formatos adecuados para el uso en LLMs.
- Desarrollar habilidades para implementar y utilizar Sentence Transformers en la creación de bases de datos vectoriales para aplicaciones RAG.
- Implementar y evaluar estrategias de búsqueda y recuperación utilizando consultas vectoriales y modelos LLM.
- Implementar y desplegar modelos LLM en entornos locales y en la nube, utilizando herramientas avanzadas para facilitar el acceso y la interacción.
- Utilizar frameworks de código abierto para desarrollar y desplegar aplicaciones RAG, demostrando comprensión y habilidad en la integración de estos componentes

Contenidos

1.-Conceptos básicos sobre LLMs y recuperación de datos



- Componentes clave de los modelos de lenguaje grande (LLMs)
 - ¿Pero qué es en esencia un LLM y qué es capaz de hacer?
 - ¿Por qué son tan revolucionarios los Transformers?
- Limitaciones de las redes neuronales clásicas
 - Redes Neuronales Convolucionales (CNNs)
 - Redes Neuronales Recurrentes (RNNs)
- La revolución de los Transformers
 - Mecanismo de autoatención
 - Atención múltiple
 - Capacidad para capturar dependencias a largo plazo
 - No secuencialidad en el entrenamiento
- ¿Y a efectos prácticos, qué necesito tener para operar mi propio LLM?
 - Lo que necesitamos saber para empezar a programar nuestro LLM
 - ¿Qué es Hugging Face?
 - ¿Qué es Google Colab?
- El Mecanismo de autoatención de nuevo: comprender relaciones entre elementos
- Representación de datos: Todo es una secuencia
- Multimodalidad: Integración de diferentes fuentes de información
- Escalabilidad y paralelización
- Entrenamiento en grandes cantidades de datos
- Cargando un modelo de GPT2 en nuestro entorno de desarrollo
- Las matemáticas muy resumidas detrás de los Transformers (solo para fans verdaderos)
- Autoatención
 - Cálculo de la salida en la capa de autoatención
 - Atención múltiple
 - Capa de proyección y FFN
 - Resumen gráfico de flujo
- Comparación de Transformers con CNNs y RNNs
- ¿Y cómo es posible que los LLMs parezcan razonar?
- Escala y profundidad del entrenamiento
- Atención y manejo de contextos largos
- Instrucciones y prompts
- Capacidades emergentes
- Razonamiento implícito mediante datos
- Componentes clave de la recuperación de datos
 - Definición y funcionalidad de RAG
 - Beneficios y aplicaciones de RAG
 - Ejemplos de aplicación de RAG
 - Fases del proceso RAG
- Resumen
 - ¿Qué son los modelos LLMs?
 - Transformers: La clave del avance
 - Funcionalidad de los LLMs
 - Ventajas de los Transformers frente a modelos clásicos



Componentes clave de los Transformers

2.-Procesamiento de datos no estructurados

- Procesamiento de datos no estructurados

Definición de datos no estructurados

Algunos tipos de datos no estructurados

¿Cómo se usan los datos no estructurados en RAG?

Preprocesamiento de Texto

¿Cuáles serían los siguientes pasos?

- Datos en ficheros

Instalar las librerías necesarias para la extracción de los datos

Funciones para leer diferentes tipos de archivos VIDEO

Procesar un repositorio de documentos

Generar embeddings para cada documento y guardarlos para su uso posterior

Cargar los embeddings guardados

- Importancia de procesar datos no estructurados para LLMs

- ¿Qué es exactamente la indexación de datos?

Limpieza y normalización de los datos

Dividir los datos en fragmentos (chunking)

Representación de los datos (Vectorización)

Construcción del índice

Consulta y recuperación

Actualización del índice

- Procesamiento de datos multimedia

Flujo para procesar multimedia (videos e imágenes)

- Procesar imágenes (extraer texto con OCR + generar embeddings visuales)

Procesar videos (extraer audio, transcribirlo y generar embeddings)

Instalación de librerías necesarias

Código para procesar videos

- Resumen

Procesamiento de datos no estructurados

Preprocesamiento de datos no estructurados en RAG

3.-Embeddings y bases de datos vectoriales

- Opciones de almacenamiento para embeddings

Bases de datos de vectores

Almacenamiento en bases de datos relacionales y NoSQL

Archivos binarios y almacenamiento en disco

Bases de datos vectoriales

¿Qué son las bases de datos vectoriales?

¿Cómo funcionan?

¿Cómo se comparan las bases de datos vectoriales?

Almacenamiento de embeddings en bases de datos vectoriales

Introducción a los Sentence Transformers

Flujo general de integración de datos procesados en LLMs

Flujo de trabajo para el Fine-tuning



Resumen

¿Cómo guardar los embeddings?

Bases de datos vectoriales

Almacenamiento de embeddings

Flujo de integración de datos en LLMs

Fine-tuning del modelo

4.-Crea tu chatbot con datos personalizados

- Creación del chatbot personalizado

Siguientes pasos

- ¿Puedo alojar mi chatbot en Hugging Face?

Crear una cuenta en Hugging Face

Crear un nuevo Space

Subir tu código del chatbot

Especificar dependencias

Probar y lanzar tu chatbot

Compartir tu Space

Ejemplo de estructura de archivos

- Alojando el chatbot en tu propio servidor

Configurar el Servidor

Instalar dependencias

- Mantenimiento del chatbot

Monitorización del chatbot

- Evaluación de la recuperación de información

Métricas relevantes

Evaluando la efectividad del chatbot

Instalación de evaluate

Ejemplo de Cálculo de Recall y MRR

- Resumen

Recuperando Información

Integración del chatbot personalizado

Evaluación de estrategias de búsqueda

5.-Despliegues de LLMs, plataformas y herramientas

- Introducción a los despliegues de LLMs

Beneficios del despliegue local versus en la nube

Casos de uso de despliegues de LLMs

Consideraciones previas para el despliegue, hardware, memoria, capacidad de procesamiento

- Herramientas y entornos para despliegue local

Introducción a entornos y plataformas

Configuración de un entorno local para desplegar LLMs

Comparación entre frameworks y herramientas

- Despliegue de LLMs en la nube

Creación y configuración de una instancia en la nube para alojar LLMs

Servicios especializados para despliegues en la nube (Amazon Sagemaker, Google Vertex IA...

- Optimización y mantenimiento de despliegues



Métodos para optimizar: cuantización podado, distillation

Consideraciones de seguridad en los despliegues de LLMs

- Estrategias de actualización y mantenimiento de los modelos

Discusión sobre desafíos y soluciones en el uso práctico de LLMs

- Resumen

Despliegue de LLMs: Local vs. nube

Casos de uso

Consideraciones del despliegue

Herramientas y entornos

Optimización y desafíos

Tal vez te interese este curso: [n8n desde cero: automatización de flujos y agentes IA sin código Nuevo](#)

O quizá este otro: [ChatGPT para desarrolladores: desde la interacción hasta la integración](#)

Síguenos en: [Instagram](#)

