

CONSIGNA DEL TRABAJO FINAL: PROYECTO FINAL

LOGRO DE APRENDIZAJE: Al finalizar el curso, el estudiante desarrolla aplicaciones innovadoras utilizando herramientas para el manejo de LLMs y técnicas de procesamiento de lenguaje natural.

1. Descripción de la actividad

Tú y 4 compañeros más colaboran en la compañía ABC, encargados de diseñar un asistente virtual avanzado para potenciar el servicio al cliente y agilizar las operaciones de atención en línea. El objetivo principal es brindar respuestas instantáneas y precisas a las consultas de los clientes, además de ofrecer recomendaciones personalizadas sobre productos o servicios. Para lograrlo, empleamos modelos de lenguaje de gran escala (LLMs) para comprender el contexto de las consultas de los usuarios y generar respuestas coherentes y pertinentes basadas en información interna relevante, como reglamentos, términos y condiciones, y detalles de productos o servicios. Nuestra meta es desarrollar un asistente virtual basado en RAG capaz de mantener diálogos naturales y efectivos con los clientes, asegurando la precisión y veracidad de la información proporcionada. Este proyecto implica el diseño, desarrollo e implementación del asistente virtual, su integración en los canales de atención al cliente de la compañía ABC, y la evaluación de su desempeño mediante métricas específicas de satisfacción del cliente y eficiencia operativa.

2. Lineamientos generales

HACKATHON 1: Desarrollo de una aplicación Q & A básica utilizando Python.

Herramientas: Python, API LLM

HACKATHON 2: Desarrollo de una aplicación Q & A optimizada con Prompt Engineering.

Herramientas: Python, API LLM, Prompt Engineering.

HACKATHON 3: Desarrollo de una aplicación Q & A con contexto aumentado utilizando Embeddings.

Herramientas: Python, API LLM, Prompt Engineering, Embeddings.

HACKATHON 4 (Avance del proyecto): Desarrollo de una aplicación Q & A con conocimiento de un caso de negocio.

HACKATHON 5: Desarrollo de una aplicación Chatbot con memoria y contexto del caso de negocio (utilizando frameworks y base de datos vectoriales).

Herramientas: Python, API LLM, Prompt Engineering, Embeddings, LangChain, LlamaIndex, BDV (ChromaDB o Pinecone), Streamlit.

HACKATHON 6: Desarrollo de una aplicación Chatbot que implementa RAG, el cual se enfoca en una necesidad específica del caso de negocio.

Herramientas: Python, API LLM, Prompt Engineering, Embeddings, LangChain, LlamaIndex, BDV (ChromaDB o Pinecone), Streamlit.

HACKATHON 7: Aplica consideraciones éticas en su proyecto el cual será sometido a pruebas éticas (ejemplo: datos personales, información falsa).

Herramientas: Python, API LLM, Prompt Engineering, Embeddings, LangChain, LlamaIndex, BDV (ChromaDB o Pinecone), Streamlit.

HACKATHON 8 (Entrega de proyecto final): Desarrollo de un Agente RAG con un profundo conocimiento del caso de negocio

3. Entregables:

- **Avance del Proyecto Final:** a la mitad del curso, el estudiante entrega el diseño preliminar del asistente virtual (modelo Q & A) según el caso del negocio, utilizando herramientas como Python, API LLM, Prompt Engineering, Embeddings, Streamlit según sea necesario.

- **Proyecto final:** al finalizar el curso, el estudiante entrega un asistente virtual avanzado basado en el modelo RAG para mejorar el servicio al cliente y optimizar las operaciones de atención en línea en la compañía ABC.

4. Presentación

- **Presentación Grupal** en Power Point o herramienta equivalente que explique el problema y la solución. Dicha presentación se mostrará antes de la demostración en vivo del funcionamiento del sistema

Repositorio:

- Buenas prácticas de documentación.
- Alojado en GitHub o Collab privado compartido con el docente.
- Incluir enlaces activos y referencias a los recursos utilizados.
- Características adicionales:
 - Especificar los lenguajes de programación utilizados.
 - Detallar las bibliotecas o frameworks empleados.
 - Mencionar la compatibilidad con SO específicos, si corresponde.

Sustentación:

Tiempo: Se asigna un total de 15 minutos para la presentación.

Actividad específica para demostrar: Durante la presentación, se espera que el equipo realice una demostración en vivo de las piezas gráficas y/o el funcionamiento del aplicativo desarrollado. Esta demostración puede incluir la navegación por la interfaz del prototipo, la presentación de ejemplos de respuestas generadas por el modelo y la explicación de cómo resuelve la necesidad del negocio.

5. Anexos

- LlamaIndex Team (2023). LlamaIndex v0.10.18.post1 [Software]. LlamaIndex: llamaindex.ai.
- Chase, H. (2022). Langchain: langchain.org.
- Huber, J. & Troynikov, A. (2022). Chroma: trychroma.com.
- (2024). Pinecone.io: pinecone.io.
- Dwight, J. (2024). Retrieval Augmented Generation (RAG) for LLMs: Your fast guide to creating accurate high-quality texts (large language models) (RAG for LLMs: Your Fast Track to Perfect Text).
- Bird, S. (2009). Natural Language Processing with Python. Recuperado de <https://tjzhifei.github.io/resources/NLTK.pdf>
- LlamaIndex (2024). rags: Build ChatGPT over your data, all with natural language. [Repositorio GitHub]. <https://github.com/run-llama/rags>

- LlamaIndex (2024). awesome-rag. [Repositorio GitHub]. <https://github.com/run-llama/awesome-rag>
- Platzi. (2023). Repositorio del proyecto del Curso de LangChain para Manejo y Recuperación de Documentos. [Repositorio GitHub]. <https://github.com/platzi/curso-langchain-qa-documents>
- Platzi. (2023). Repositorio del proyecto y ejercicios del Curso de LangChain. [Repositorio GitHub]. <https://github.com/platzi/curso-langchain>
- starmorph. (2023). pdf-analyze-streamlit. [Repositorio GitHub]. <https://github.com/starmorph/pdf-analyze-streamlit>
- López, R.E. (s.f.) Introducción a Python. <https://iaarbook.github.io/python/>
- López, R.E. (s.f.) Introducción al Machine Learning. <https://iaarbook.github.io/ML/>
- López, R.E. (s.f.) Introducción a la Big Data. <https://iaarbook.github.io/bigdata/>