

CHATBOT PARA ATENCIÓN AL CLIENTE

Justificación

Se propuso un chatbot en el área de servicio al cliente porque permite atender a los usuarios de forma inmediata, sin depender exclusivamente de la disponibilidad de agentes humanos. Esto mejora drásticamente los tiempos de respuesta, y es muy útil en horarios no laborables o durante picos de alta demanda.

Además, los chatbots pueden resolver consultas repetitivas y sencillas permitiendo que los agentes humanos se concentren en casos más complejos que requieren mayor análisis.

También contribuyen a la reducción de costos operativos y aseguran una experiencia de atención consistente, lo que incrementa la satisfacción del cliente.

En este caso se había propuesto la aplicación de un chatbot para responder preguntas relacionadas con documentación interna de una empresa, sin embargo por políticas internas no se otorgó el acceso a esta información, por lo que se hizo el proyecto basándose en los términos y condiciones de una aerolínea mexicana. Esa información es de carácter público. Se mantuvo el objetivo de hacer un chatbot orientado a resolver preguntas de los clientes.

Construcción del código

El chatbot utiliza tres modelos clave:

El modelo de embeddings **all-MiniLM-L6-v2**, se usa para convertir tanto los fragmentos del documento como la pregunta del usuario en vectores numéricos. Este modelo es ideal para tareas de similitud semántica donde se necesita comparar textos de manera eficiente.

Para ordenar los fragmentos más relevantes, se utiliza el modelo **cross-encoder/ms-marco-MiniLM-L-6-v2**, que evalúa cada pregunta con su posible respuesta y les asigna una puntuación de relevancia, a diferencia del modelo de embeddings que compara de forma "independiente".

Finalmente, se usa el modelo de lenguaje **tiiuae/falcon-7b-instruct** para seguir instrucciones. Se encarga de generar la respuesta final en lenguaje natural a partir del contexto. Este modelo tiene la capacidad de generar respuestas coherentes con el contenido proporcionado.

Esta combinación de modelos permite la comprensión y generación de texto: primero se identifica el contenido relevante, y luego se genera una respuesta basada en ese contenido.