

Procesamiento de Lenguaje Natural

Maestría en Inteligencia Artificial Aplicada Tecnológico de Monterrey

Tarea: Chatbot con RAG+LLM

Número de Equipo: _____

Nombres: _____

Matrículas: _____

La presente actividad se realiza en equipos.

Cada equipo diseñará e implementará un chatbot basado en un modelo de lenguaje (LLM) con un sistema de recuperación aumentada generativa (RAG). El chatbot debe abordar una temática o problemática específica elegida por el equipo y demostrar cómo la combinación de LLM y RAG optimiza la interacción y generación de respuestas, además de personalizarlas a la temática en cuestión.

1. Cada equipo tendrá plena autonomía para definir la temática del chatbot.
2. Los equipos podrán elegir el modelo de lenguaje de gran tamaño, LLM, que consideren más adecuado para su chatbot. Podrán optar por modelos open-source o propietarios.
3. Integrar un sistema de recuperación aumentada generativa (RAG) que permita mejorar la precisión de las respuestas mediante la consulta de documentos relevantes, seleccionados o generados por el equipo.
4. Se recomienda utilizar documentos en español para alimentar el sistema RAG y garantizar que la salida generada por el chatbot también esté en español. Sin embargo, puede ser en otro idioma si la problemática así lo requiere.
5. Incluir una introducción y justificación donde se describa la temática elegida y la problemática que el chatbot busca resolver o apoyar.
6. El chatbot puede ser directamente en el cuaderno de Jupyter, pero se puede enriquecer con alguna interfaz gráfica como Gradio.
7. Los equipos deberán entregar un informe documentando de la manera en que funciona el chatbot con un sistema RAG+LLM.
8. Incluyan ejemplos de varias preguntas y las respuestas que les genera el sistema.
9. Incluyan sus conclusiones finales de la actividad, en particular en relación a qué tan adecuadas consideran las respuestas en relación al contexto de la temática considerada.