

Tarea: Evaluación y optimización de un sistema RAG con Llama3 (usar modelos alternativos a Llama3) (o similares) y ChromaDB

Objetivo: Modificar y evaluar el comportamiento del sistema al realizar consultas con y sin RAG, analizando los impactos en relevancia, ética y claridad.

Indicaciones:

1. Cambios en los parámetros:

- Modificar el valor de `n_results` en la función `collection.query()` a 1 y 5.
- Ejecutar pruebas con cada valor para observar cómo afecta la relevancia y claridad de las respuestas.
- Comparar respuestas al deshabilitar RAG (es decir, enviar el prompt directamente al modelo sin incluir contexto desde ChromaDB).

2. Escenarios de prueba: Seleccionar 3 preguntas de escenarios reales relevantes basados en el contexto que se va a cargar a ChromaDB (`contexto.txt`).

3. Evaluación: Crear una tabla para comparar las respuestas obtenidas con y sin RAG y con los diferentes `n_results` en las siguientes categorías:

- Relevancia: ¿La respuesta se alinea con el contexto?
- Claridad: ¿La respuesta es comprensible y directa?
- Ética: ¿La respuesta refleja principios éticos?

Utilizar la siguiente escala:

- 0: Insatisfactorio.
- 1: Parcialmente satisfactorio.
- 2: Satisfactorio.

4. Análisis de Resultados:

- Calcular la puntuación total para cada categoría y método (con RAG vs. sin RAG y `n_results` 1 vs 5).
- Responder a las siguientes preguntas:
 1. ¿El uso de RAG mejora la relevancia y claridad?
 2. ¿Qué limitaciones tiene el sistema actual con RAG?

5. Representación gráfica: Crear gráficos de barras para comparar las puntuaciones totales entre los diferentes valores de `n_results` y entre el uso de RAG vs. sin RAG.

6. Documentación y evidencias:

- Realiza un informe breve (máximo 1000 palabras).
- Coloca principales hallazgos de la evaluación.
- Aborde las limitaciones observadas en el sistema.
- Sugerencias para optimizar el sistema en el futuro.
- Capturas de pantalla de las respuestas obtenidas para cada caso.
- Agrega capturas las partes del código donde realizaron aportaciones propias.
- El Informe no debe exceder las 1000 palabras.
- Redacta una breve conclusión de máximo 150 palabras interpretando los resultados obtenidos.

Ejemplo de la tabla:

Pregunta	RAG?	# n_results	Relevancia	Claridad	Ética
¿... ?	Si / No	1 / 5	0-2	0-2	0-2
¿... ?					
¿... ?					