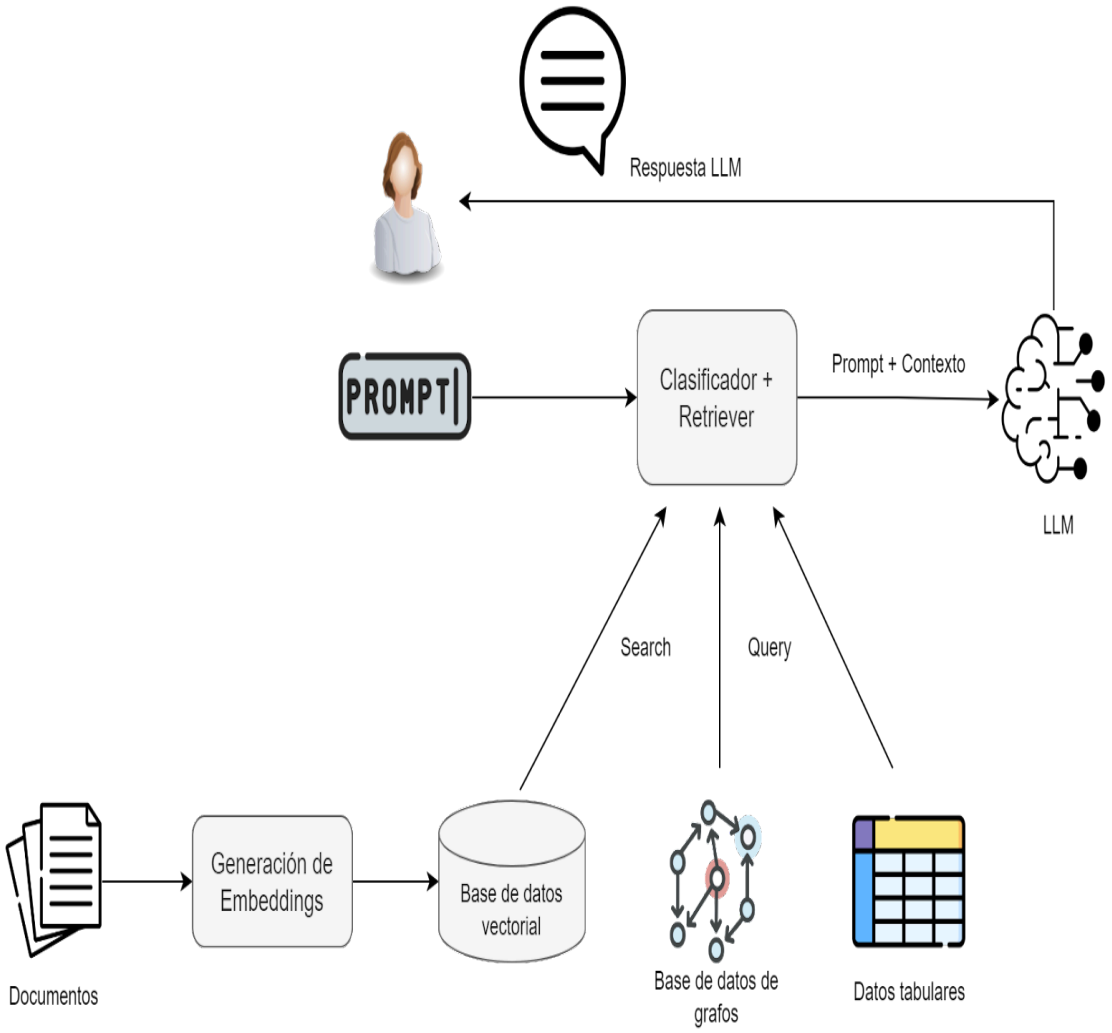

<u>Estudiante</u>: Ponce, Daniel

Antes de comenzar esta versión del trabajo práctico, surgió de las correcciones sugeridas por la cátedra con el fin de mejorar el trabajo práctico, por lo cual se agradecen las sugerencias y que me permitieron tener una mejor versión de dicho archivo.

PUNTO 1

El tema elegido para el chatbot, fue que se un experto en ciberataques debido a esto las fuentes de datos que elegí fueron las siguientes seguinfo, cve mitre y exploit db.

En los comienzo esta idea el enfoque por el lado que sólo respondiera que era un exploit pero cuando vi realmente el potencia que le brindaba usar rag con los LLM, el enfoque fue cambiando y una idea se asomo que fue porque no sumar un base de datos de exploits.

No todo fue color de rosas, hubo muchos problemas que tuve que solucionar, como en este caso la idea de hacer el scrapper al principio funcionaba muy bien, luego me tope con la limitación de las peticiones al servidor y los bloqueos.

Con ello me llevó a implementar proxys chains , luego me bloquearon la ip y todos los proxys estaban baneados y a razón de eso tuve que cambiar de colab.

Luego me dediqué un tiempo a buscar un base de datos de grafos para este proyecto, pero no conseguí alguna que sirviera para lo que era mi idea. Con estas negativas decidí crear una base de datos propia usando neo4j.

En dicha base de datos use el archivo de exploit db y represente cada exploit como un nodos, y cree las relaciones **afecta_a** que es la manera de relacionar un platform con los exploits.

Pero este proceso lleva mucho tiempo, y con objetivo de hacerlo solo para este TP reducir los datos solo las vulnerabilidades a las que parecieron a partir del 2018 en adelante.

En el caso de los datos obtenidos con el scrapper y la base de datos de cve mitre hice los mismo decidí achicar el tamaño de ambos para obtener mejor tiempo en la construcción de los embedding.

Previamente había experimentado con procesamiento en paralelo y con múltiples procesos pero no tuve un decremento importante considerable en base a esto decidí tomar la decisión que conté con anterioridad.

A modo conclusión puedo decir, que este trabajo me dejo muchas enseñanzas entre las cuales puedo destacar el aprender a crear embedding, y aprender cómo trabajar con un modelo llm remoto, y comprender cómo funcionaba los Rag que no

los conocía. En comparación con el chatbot que cree en el primer trabajo este es mucho más humano y no tan artificial.