

## 1. Descripción técnica-conceptual del proyecto a realizar

El objetivo de este proyecto es desarrollar un chatbot especializado que pueda ser entrenado con documentos propietarios de una empresa, y que, en base al contenido ingestado, interprete correctamente consultas de usuarios y proporcione respuestas precisas y relevantes.

En un entorno empresarial, la eficiencia en la búsqueda de información es crucial para la productividad y el rendimiento de los empleados. Sin embargo, con la creciente cantidad de datos y documentos disponibles, encontrar información específica de manera rápida y precisa puede convertirse en un desafío. Un chatbot especializado ofrece una solución prometedora al permitir a los usuarios realizar consultas en lenguaje natural y obtener respuestas de manera instantánea.

Mientras que otros sistemas de inteligencia artificial ampliamente conocidos y utilizados, como ChatGPT o Microsoft Copilot, destacan en su capacidad para generar respuestas generales basadas en un amplio conocimiento del lenguaje, el proyecto propuesto se distingue por su capacidad para trabajar con documentos altamente específicos. Esto le permitirá ofrecer respuestas adaptadas al contexto interno de la organización, las cuales no podrían obtenerse mediante el uso de los chatbots de propósito general disponibles en el mercado.

En la figura 1 se presenta el diagrama en bloques del sistema. Se observa que en primera instancia, los documentos son procesados y divididos en fragmentos de texto, que luego se convierten en vectores matemáticos llamados *embeddings*. Estos vectores son una representación del significado subyacente de las palabras, lo que permite que las computadoras procesen el lenguaje de manera más efectiva. Se utiliza una base de datos vectorial para almacenar la información procesada. Por otro lado, los usuarios interactúan con el chatbot a través de una interfaz gráfica, donde consultan por la información deseada. Estas consultas son luego utilizadas por un proceso de búsqueda por similitud para identificar los fragmentos más relevantes. Estos fragmentos, junto con la consulta del usuario, se proporcionan a un modelo de lenguaje de gran tamaño (LLM), que genera una respuesta coherente y contextualizada. Finalmente, la respuesta se presenta al usuario en la interfaz gráfica, completando el ciclo de interacción.

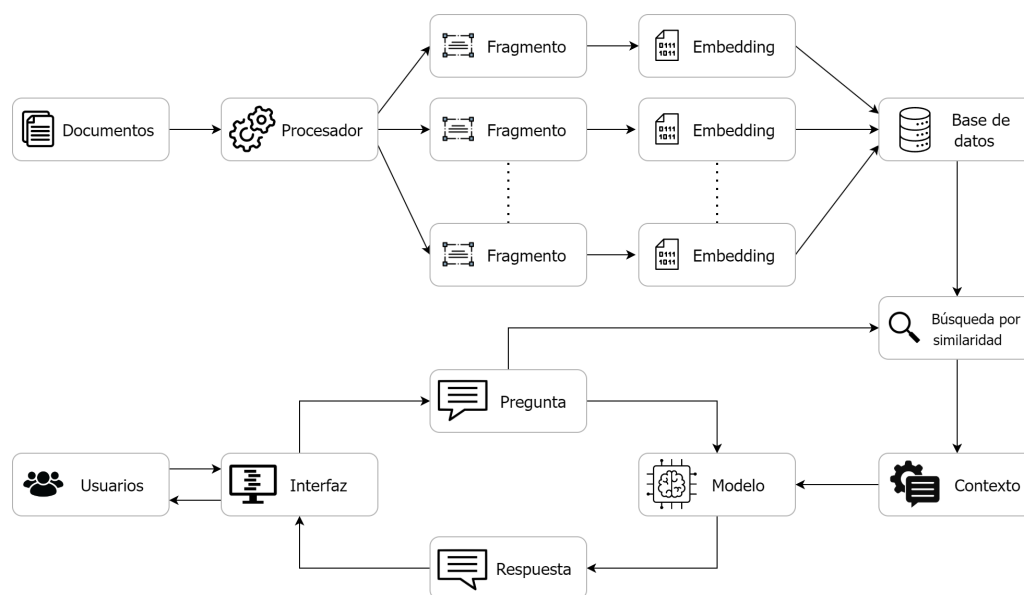


Figura 1. Diagrama en bloques del sistema.