El desarrollo del proyecto `llm\_chatbot` permitió comprender de manera práctica cómo integrar distintos modelos de lenguaje (OpenRouter, Google Gemini y Ollama) en una misma aplicación utilizando la librería `OpenAI` como interfaz común. A través del uso de Gradio, se construyó una interfaz gráfica sencilla e interactiva que facilita la comunicación con los modelos y la ejecución de tareas como traducción y resumen. Este proceso fortaleció conocimientos sobre modularidad del código, gestión de claves mediante `dotenv`, y adaptación de APIs con diferentes estructuras de respuesta.

Durante la implementación se identificaron algunos límites técnicos, como la falta de estandarización entre las respuestas de los modelos, errores genéricos poco informativos y la ausencia de persistencia del historial de conversación, que se pierde al cerrar la aplicación. Además, se evidenciaron restricciones en el uso de modelos gratuitos, con tiempos de respuesta variables y limitaciones de uso.

Entre las posibles mejoras destacan la incorporación de un sistema de almacenamiento de sesiones para conservar conversaciones, el manejo más detallado de errores, la ampliación de los tipos de tareas del chatbot y la optimización de la interfaz con funciones adicionales. También se plantea la posibilidad de desplegar el sistema en la nube mediante plataformas como Gradio Spaces o Streamlit Cloud. En conjunto, este proyecto aportó una visión integral sobre la integración de LLMs, el diseño modular y la creación de herramientas escalables basadas en inteligencia artificial.