

Kenneth Eliécer Rojas Rivera

NAO ID: 3477

19/12/2025

Integración de LLMs en programación de APIs

Sprint 2

Historias de usuarios

Historia 1 — Procesamiento de documentos

Como investigador universitario quiero subir archivos .pdf, .txt y .html al asistente para lograr obtener análisis y respuestas basadas en su contenido.

Historia 2 — Respuestas generadas por LLM

Como estudiante quiero que CueBot procese mis preguntas utilizando un LLM confiable para lograr obtener explicaciones claras y útiles sobre mis documentos.

Historia 3 — API funcional

Como miembro del departamento TI quiero una API en FastAPI que permita enviar texto y recibir la respuesta del LLM para lograr integrar CueBot fácilmente con otros sistemas de la universidad.

Historia 4 — Interfaz web de prueba

Como docente quiero una interfaz web simple hecha en Gradio para lograr interactuar con CueBot sin instalar software adicional.

Historia 5 — Registro de logs

Como administrador del proyecto quiero que el sistema registre peticiones, errores y respuestas para lograr evaluar el funcionamiento del MVP y saber qué mejorar.

Tablas de registro

Tabla 1. Lista de requerimientos

Historia de usuario	Requerimientos
Procesamiento de documentos	<ul style="list-style-type: none">- Cargar archivos .pdf, .txt, .html- Extraer texto de cada formato- Enviar texto procesado al LLM
Respuestas generadas por LLM	<ul style="list-style-type: none">- Configurar acceso al modelo LLM- Implementar prompt adecuado- Recibir y mostrar respuesta
API funcional (FastAPI)	<ul style="list-style-type: none">- Crear endpoints /upload y /ask- Validación de archivos- Manejo de errores- Conexión con el LLM
Interfaz web Gradio	<ul style="list-style-type: none">- Crear interfaz con campo de texto y carga de archivos- Conectar la interfaz con la API- Mostrar resultados y errores
Registro de logs	<ul style="list-style-type: none">- Registrar peticiones en un archivo o base de datos- Registrar errores y eventos- Acceso para revisión

Tabla 2. Lista priorizada

Requerimiento	Etapas	Estimación de tiempo	Entregables
Cargar archivos (.pdf/.txt/.html)	1	6h	Función que recibe archivos y valida formato
Extraer texto de archivos	1	8h	Módulo extractor de texto funcionando
Conectar texto extraído con el LLM	1	6h	Script que envía texto al LLM y recibe respuesta
Configurar acceso al modelo LLM	1	3h	Config de API key + cliente LLM
Implementar endpoints FastAPI	2	10h	API funcional /upload y /ask
Manejo de errores en API	2	4h	Validaciones y respuestas estándar

Crear interfaz Gradio	2	8h	UI funcional para pruebas
Conectar Gradio con FastAPI	2	4h	Interacción entre UI y backend
Registrar logs (errores y peticiones)	3	6h	Sistema de logs y archivo .log
Revisar API para integración	3	6h	Ajustes de formato + documentación