# Bot de Preguntas y Respuestas sobre Convocatoria de Estancias/Estadías y Reglamento Escolar

## Descripción general

En este proyecto desarrollarás, en equipo, un bot de consulta que permita a estudiantes y docentes realizar preguntas sobre:

- La Convocatoria de Estancias y Estadías
- El Reglamento Escolar

El sistema deberá permitir:

1. Subir documentos (PDF oficiales) a través de una interfaz.
2. Procesar y almacenar estos documentos en una base de datos vectorial, de modo que el sistema pueda "entender" su contenido.
3. Responder preguntas en lenguaje natural usando un modelo de lenguaje (LLM), apoyado en la información almacenada en la base de datos vectorial (enfoque tipo *Retrieval-Augmented Generation, RAG*).
4. Administrar los documentos: listar, agregar y eliminar documentos. Si un documento se elimina de la gestión, su contenido debe eliminarse también de la base de datos vectorial para que deje de ser utilizado en las respuestas.

El objetivo es simular un asistente institucional que ayude a los estudiantes a resolver dudas frecuentes (requisitos, plazos, sanciones, trámites, etc.) de forma rápida y precisa.

## Objetivo general

Diseñar e implementar un sistema de preguntas y respuestas basado en documentos institucionales, utilizando una arquitectura con base de datos vectorial + LLM, que permita consultar y administrar de forma dinámica los documentos fuente.

## Alcance funcional mínimo (MVP)

Tu sistema debe cumplir al menos con lo siguiente:

### 1. Módulo de documentos

- **Subir documentos**:
  - Interfaz para cargar archivos (mínimo PDF).
  - Guardar metadatos básicos: nombre, tipo de documento, fecha de carga, etiqueta (por ejemplo: "Convocatoria 2025", "Reglamento Escolar 2024").
- **Procesamiento**:

    - Extraer el texto del documento.
    - Dividir en fragmentos (chunks) apropiados.
    - Generar **vectores (embeddings)** de cada fragmento.
    - Almacenar embeddings y metadatos en una **base de datos vectorial**.
- **Administración**:
    - Listar documentos cargados.
    - Eliminar documentos.
    - Al eliminar un documento:
        - Se debe eliminar el archivo del sistema.
        - Se deben eliminar también todos los vectores asociados en la base de datos vectorial (el contenido ya no puede aparecer en futuras respuestas).

**2. Módulo de preguntas y respuestas (Bot)**

- Caja de texto para que el usuario escriba una pregunta en lenguaje natural.
- Flujo de respuesta:
    1. Tomar la pregunta del usuario.
    2. Consultar la base de datos vectorial para recuperar los fragmentos de texto más relevantes.
    3. Enviar la pregunta + fragmentos relevantes al LLM.
    4. Devolver al usuario una respuesta redactada, clara y coherente con los documentos.
- Mostrar, de forma sencilla, de qué documento(s) se obtuvo la información (por ejemplo, "Fuente: Reglamento Escolar, Artículo 15").

## Entregables del proyecto

1. **Repositorio de código** (GitHub o similar) con:
    - Código fuente documentado.
    - Archivo de requisitos (requirements.txt, pyproject.toml, package.json, etc.).
    - Instrucciones de instalación y ejecución en un README.
2. **Documento técnico (máx. 10 páginas)** que incluya:
    - Descripción del problema y motivación.
    - Arquitectura del sistema (diagrama de bloques o flujo).
    - Descripción del flujo de datos.
    - Decisiones de diseño (elección de herramientas, modelos, parámetros).
    - Limitaciones y posibles mejoras futuras.
3. **Manual de usuario** (puede ser sección del mismo documento técnico) con:
    - Cómo subir documentos.
    - Cómo hacer preguntas.
    - Cómo administrar documentos (eliminar, actualizar).
4. **Presentación final**:
    - Explicación breve de la arquitectura.
    - Demo en vivo del sistema:

- Subir o mostrar documentos ya cargados.
- Realizar varias preguntas típicas.
- Eliminar un documento y mostrar que la información deja de utilizarse.