

# Asistente Académico con RAG

*Retrieval-Augmented Generation*

Iman Noriega Melissa  
Méndez Gonzalo Miguel  
Orrego Torrejón Diego  
Pineda García Diego  
Trujillo Serva Luis Andre

Universidad Nacional de Ingeniería  
Facultad de Ciencias  
Escuela Profesional de Ciencia de la Computación

Septiembre 2025

# Agenda

- 1 Descripción del Problema
- 2 Solución Propuesta
- 3 Objetivos Preliminares
- 4 Técnicas de IA Identificadas
- 5 Consideraciones Éticas
- 6 Próximos Pasos

# Contexto Actual

- Los estudiantes enfrentan dificultades para encontrar información específica en documentos académicos extensos
- Pérdida de tiempo navegando por múltiples PDFs, libros y apuntes de curso
- Necesidad de respuestas rápidas y precisas durante el estudio
- Fragmentación de la información académica en diferentes formatos

## Problema Principal

Los estudiantes pierden tiempo valioso buscando información específica en PDFs de cursos, afectando su eficiencia de estudio y comprensión.



## Solución

Un chatbot inteligente que carga apuntes, libros y documentos de un curso específico para responder preguntas de manera contextual usando técnicas de Retrieval-Augmented Generation.

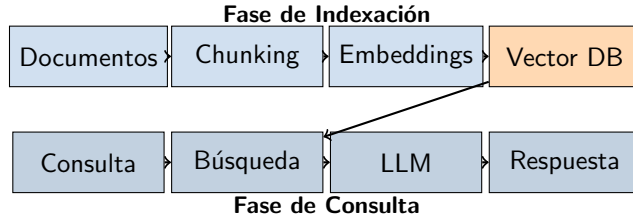
### Características principales:

- Carga y procesa documentos PDF
- Búsqueda semántica inteligente
- Respuestas contextualmente relevantes
- Interfaz conversacional natural

### Beneficios esperados:

- Acceso rápido a información
- Respuestas precisas y contextuales
- Mejora en la comprensión
- Optimización del tiempo de estudio

# Arquitectura del Sistema RAG



## 1. Indexación:

- Procesamiento de documentos
- Vectorización con embeddings
- Almacenamiento en FAISS/ChromaDB

## 2. Consulta:

- Búsqueda de similitud semántica
- Contexto relevante para LLM
- Generación de respuesta aumentada

# Objetivo General

## Objetivo General

Desarrollar un asistente académico inteligente que utilice técnicas de Retrieval-Augmented Generation para proporcionar respuestas precisas y contextualmente relevantes a partir de documentos académicos específicos de un curso.

# Objetivos Específicos

- ➊ **Objetivo Técnico:** Implementar un sistema RAG que combine embeddings vectoriales con modelos de lenguaje grande para la recuperación y generación de respuestas.
- ➋ **Objetivo de Evaluación:** Medir la precisión de las respuestas generadas usando métricas como BLEU, ROUGE y evaluación humana de relevancia contextual.
- ➌ **Objetivo de Aplicación:** Crear una herramienta práctica que reduzca el tiempo de búsqueda de información académica en al menos un 60 % comparado con métodos tradicionales.
- ➍ **Objetivo Ético:** Implementar mecanismos para identificar y mitigar sesgos en las respuestas generadas, así como respetar la propiedad intelectual de los documentos fuente.



## 1. Procesamiento de Lenguaje Natural

- Modelos de Embeddings (OpenAI, Sentence-BERT)
- Large Language Models (GPT-3.5/4, LLaMA)
- Tokenización y chunking inteligente

## 2. Búsqueda Vectorial

- FAISS (Facebook AI Similarity Search)
- ChromaDB para almacenamiento vectorial
- Similarity search con cosine distance

## 3. Arquitectura RAG

- Retrieval-Augmented Generation
- Prompt engineering
- Context window optimization

## 4. Técnicas de Evaluación

- Métricas de similitud semántica
- Evaluación de coherencia
- Análisis de relevancia contextual

## Frameworks y Librerías:

- **LangChain:** Orquestación de LLMs
- **OpenAI API:** Embeddings y GPT
- **FAISS/ChromaDB:** Vector storage
- **Streamlit:** Interfaz de usuario
- **PyPDF2:** Procesamiento de PDFs

## Justificación Teórica:

- **Embeddings:** Representación semántica densa
- **Attention Mechanisms:** Relevancia contextual
- **Transfer Learning:** Conocimiento preentrenado
- **Similarity Search:** Recuperación eficiente

## Fundamento del Curso

Este proyecto aplica conceptos fundamentales de IA estudiados: representación del conocimiento, búsqueda inteligente, procesamiento de lenguaje natural y sistemas basados en agentes.

# Aspectos Éticos Identificados

## 1. Sesgos en Respuestas

- Identificación de sesgos del modelo base
- Evaluación de equidad en respuestas
- Implementación de filtros de contenido

## 2. Propiedad Intelectual

- Respeto por derechos de autor
- Atribución apropiada de fuentes
- Uso educativo justificado

## 3. Transparencia del Sistema

- Explicabilidad de respuestas
- Trazabilidad de fuentes
- Confianza en la información

## 4. Privacidad de Datos

- Protección de información sensible
- Anonimización de consultas
- Almacenamiento seguro

## Compromiso Ético

Implementar un sistema transparente, justo y respetuoso que mejore el aprendizaje sin comprometer la integridad académica.

# Cronograma Preliminar

Semana	Actividad	Entregable
4-5	Investigación y diseño técnico	Marco teórico
6-7	Implementación del pipeline RAG	Prototipo básico
8-9	Desarrollo de interfaz y optimización	Sistema funcional
10-11	Evaluación y pruebas	Métricas y resultados
12-13	Análisis ético y documentación	Informe técnico
14-15	Preparación de presentación final	Presentación

## Recursos necesarios:

- API keys de OpenAI
- Documentos de prueba (PDFs académicos)
- Infraestructura de cómputo para embeddings

## Contacto del Grupo:

- Coordinador: [Luis Andre Trujillo Serva]
- Email: [luis.trujillo.s@uni.edu.pe]

*¡Gracias por su atención!*