

L U M B R E



El pipeline RAG: del documento a la respuesta con fuente

Materia: IA Aplicada, Vibe Coding y Negocios Digitales

07/10/2026

Antes de empezar

RAG no es un truco, es una cadena con pasos. Recorre el camino completo desde que subes un documento hasta que la IA responde citando de dónde salió.

QUÉ APRENDERÁS

- Seguir de principio a fin la cadena RAG (fragmentar, embeber, guardar, buscar, armar contexto, responder citando) y reconocer qué pieza falla cuando la respuesta es mala.

La cadena completa, paso a paso

Mapa conceptual

El pipeline RAG: del documento a la respuesta con fuente

La cadena completa, paso a paso

Dónde se rompen las cosas

El mapa mental interactivo es interactivo en la versión online.



Categorías a serem representadas na criação de um documento institucional



Imagen libre de Wikimedia Commons (CC BY 4.0, autor: Prosa)

VIDEO

¿Qué es RAG? - Clase con código y ejemplo

El pipeline RAG explicado con ejemplo real. Aquí verás de qué hablamos cuando decimos anclar la IA en tus propios documentos.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=uAsd9p0IcLg>

Un sistema RAG funciona como una línea de producción. Si la respuesta final es mala, casi siempre el problema está en **una etapa concreta**, no en 'la IA'. Aprender a

nombrar las etapas te permite diagnosticar y arreglar en vez de adivinar.

Hay dos momentos: **indexar** (preparar tus documentos una vez) y **consultar** (responder cada pregunta).

1. Ordena las etapas del pipeline RAG, desde que subes un documento hasta que recibes la respuesta con fuente.

- Fragmentar el documento en pedazos manejables (chunks).
- Convertir cada pedazo en un embedding (huella de significado).
- Guardar los embeddings en la base vectorial.
- Convertir la pregunta del usuario en un embedding.
- Buscar los pedazos más cercanos a la pregunta.
- Armar el contexto con esos pedazos y dárselo a la IA.
- Responder citando de qué pedazo salió cada dato.

Dónde se rompen las cosas

Diagnóstico rápido: el síntoma apunta a la etapa, y la etapa tiene un control.

Síntoma en la respuesta	Etapas probable que falla	Qué ajustar
No encuentra nada aunque el dato existe	Búsqueda / embeddings	Revisar cómo fragmentaste o el tamaño de pedazo
Responde con datos de un área equivocada	Armado de contexto	Pedir menos pedazos o filtrar por documento
Cita un pedazo pero lo interpreta mal	Generación (la IA)	Mejorar la instrucción de "responde solo con el contexto"
Inventó un dato que no estaba en ningún pedazo	Falta de anclaje	Exigir cita obligatoria o rechazar si no hay fuente

Contexto no es todo el documento

La IA tiene un **límite de contexto** (cuánto texto admite por mensaje). No le pegues el PDF completo: dile que busque y solo entrégale los pedazos relevantes. Más texto innecesario = más ruido, más costo y peores respuestas.

2. Tu asistente RAG responde bien preguntas de "Vacaciones" pero mal las de "Nómina", aunque el manual de nómina sí está subido. ¿Cuál es la causa más probable?

- ☐ La IA odia la nómina.
- ☐ El manual de nómina se fragmentó tan mal (o tan grande) que la búsqueda no trae el pedazo correcto.
- ☐ La pregunta de nómina siempre llega en otro idioma.
- ☐ No hay causa: es azar.

3. En RAG, la indexación de documentos se repite cada vez que un usuario hace una pregunta.

- ☐ Verdadero
- ☐ Falso

4. Piensa en un documento real de tu negocio (un manual, una lista de precios, unos términos). Describe en dos líneas cómo lo fragmentarías: ¿por secciones, por preguntas frecuentes, por página? ¿Qué podría salir mal si lo haces de una sola pieza gigante?

Un pedazo que mezcla tres temas confunde a la búsqueda.

Siguiente paso: Proyecto — diseñarás la arquitectura de consulta documental de una empresa real, eligiendo fuentes, fragmentación y reglas de respuesta.

Referencia curricular

Materia:	IA Aplicada, Vibe Coding y Negocios Digitales
-----------------	---

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://empresa.lumbre.lat/object/iav-c13-r01-0a03>

Material educativo generado desde Lumbre · lumbre.lat