

# Aplicaciones LLM

## Embeddings

# Recuerda: RAG supera context window

- Preparativos:
  - Divide tus datos en pequeños segmentos.
  - Convierte los pequeños segmentos en números (“embeddings”).
  - Carga los embeddings en una base de datos vectorial.
- Ahora cuando preguntas (“query”) al LLM:
  - El LLM va a la base de datos vectorial y busca (“indexing”) solo (“semantic similarity”) los datos que responden a tu pregunta.

# Embeddings

- Los ordenadores trabajan con números.
- Por eso convierten el texto en números.
- Hacen lo mismo con las imágenes, el audio, el video, etc.
- Los embeddings son vectores de números.
  - Ejemplo: “hola” se convierte en un embedding como (1,4,6).