

RAG (Retrieval-Augmented Generation) es una técnica que combina recuperación de información con generación de texto. En lugar de depender solo del conocimiento entrenado, el modelo busca contexto relevante en documentos externos y lo utiliza para generar una respuesta precisa. Se usa en chatbots, sistemas legales, atención médica y asistentes con memoria. Utiliza bases vectoriales, embeddings, y modelos como GPT para generar salidas contextualizadas y confiables...