Topic 13: Multimodal RAG

1. Guide to Multimodal RAG for Images and Text: https://medium.com/kx-systems/guide-to-multimodal-rag-for-images-and-text-10dab36e3117
2. RAG with Images and Tables: A practical tutorial on enhancing RAG to handle images, tables, and charts in documents: https://medium.com/@rajratangulab.more/building-a-multimodal-rag-chatbot-with-image-text-and-table-understanding-91946bc9c51c
3. A Comprehensive Guide to Building Multimodal RAG Systems: https://www.analyticsvidhya.com/blog/2024/09/guide-to-building-multimodal-rag-systems/
4. An Easy Introduction to Multimodal RAG: NVIDIA's technical blog explaining the challenges of multiple modalities and outlining approaches like unified embeddings, image captioning, or separate stores with a re-ranker: https://developer.nvidia.com/blog/an-easy-introduction-to-multimodal-retrieval-augmented-generation/
5. A Hugging Face community article showing a true multimodal RAG system indexing both image embeddings and text embeddings, and using a multimodal LLM for retrieval + generation: https://huggingface.co/blog/Omartificial-Intelligence-Space/building-multimodal-rag-systems
6. A tutorial from HF showing how to build a pipeline retrieving both text and images and feeding them into a visual-language model: https://huggingface.co/learn/cookbook/en/multimodal_rag_using_document_retrieval_and_vlms
7. Enhance Your RAG Agent with Multimodal Retrieval: A Databricks engineering blog showing how to add image search to a RAG system using vector search and multimodal embeddings: https://medium.com/@AI-on-Databricks/enhance-your-rag-agent-with-multimodal-retrieval-4d15b0b173de
8. Multimodal Chart Retrieval: A research paper from Google DeepMind tackling retrieval from charts/plots by converting to tables, using a vision-language model: https://arxiv.org/abs/2412.10704
9. Survey on Multimodal RAG: A systematic survey of multimodal RAG for document understanding, summarising datasets, benchmarks and open problems: https://arxiv.org/abs/2510.15253
10. A video tutorial on how to build a multimodal RAG system that chats with PDFs containing images and tables: https://www.youtube.com/watch?v=uLrReyH5cu0
11. RAG-Anything: All-in-One Multimodal RAG Framework: https://github.com/HKUDS/RAG-Anything
12. LangChain's blog post introducing multi-vector retriever cookbooks for handling documents with mixed content types: https://blog.langchain.dev/semi-structured-multi-modal-rag/