



@Joanna
Stoffregen

RAG

Ultimate cheat sheet

Definition



RAG stands for **R**etrieval-**A**ugmented-**G**eneration
In other words : your favorite **LLM** (ChatGPT)
augmented with **your data/documents** at scale

What for ?

- 1.Doc Q&A
- 2.Customer support
- 3.Process automation
- 4.Data analysis

Data layer

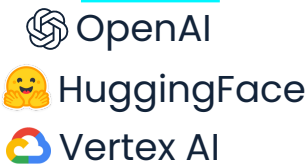


The **3 foundations of RAG** applications are

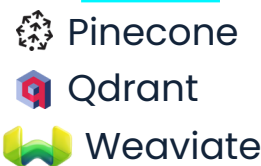
Data preparation



Embeddings



Vector storage



Enterprise RAG



Use **Azure Databricks**



Application

RAG chain

Monitoring

Models

Data & vector pipe

Lakehouse : storage & governance

RAG for beginners



GPTs + code interpreter



Copilot **NEW feature**

Private RAG



Verba by Weaviate
(open source)



NVIDIA chat with RTX



Ollama + Mistral AI

Use case examples



A lawyer who needs insight on a complex legal case
with confidential documents >> **Use Private RAG**



An accountant who needs to get immediate insights
on his data (for a few docs) >> **Use RAG for beginners**



Internal knowledge chatbot for a big firm >> **Enterprise RAG**