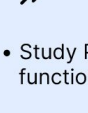


Become a RAG Specialist in 2025

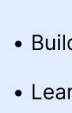
Step 1

Master the Programming Foundations



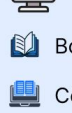
Actions

- Study Python basics: variables, loops, conditionals, and functions.
- Explore data structures: lists, dictionaries, sets, and tuples.
- Learn file handling to process JSON, CSV, and other formats.
- Understand object-oriented programming (OOP) concepts.
- Practice coding through projects and exercises.



Goals

- Build a strong grasp of Python programming.
- Learn to write modular, reusable, and efficient code.



Resources

- Books: Automate the Boring Stuff with Python.
- Courses: Analytics Vidhya beginner Python tutorials.
- Tools: Code editors like VS Code or PyCharm.

Step 2

Learn Core Libraries and Tools



Hugging Face



PyTorch



Chroma



Actions

- Familiarize yourself with ML/DL frameworks: TensorFlow, PyTorch.
- Explore NLP-specific libraries: Hugging Face.
- Learn tools like LangChain for RAG pipelines.
- Understand vector databases like ChromaDB for efficient similarity search.
- Learn how to access commercial models such as GPT-4o, Claude 3.5 & Gemini 1.5.
- These tools help streamline tasks like data preprocessing, embeddings, and retrieval.



Goals

- Become proficient with essential libraries and tools for RAG development.

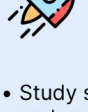
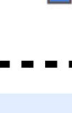


Resources

- LangChain Cookbooks
- Tutorials: Official documentation and community guides for different libraries.
- Projects: Implement toy models using these tools.

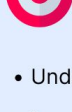
Step 3

Grasp Machine Learning and Information Retrieval Basics



Actions

- Study supervised learning (e.g., regression, classification) and unsupervised learning (e.g., clustering).
- Explore algorithms for document relevance: TF-IDF, BM25, PageRank and Cosine Similarity.
- Understand vector space models and their role in retrieval.



Goals

- Understand foundational machine learning concepts.
- Learn information retrieval techniques.



Resources

- Course: Machine Learning Certification Course for Beginners by Analytics Vidhya
- Coursera ML by Andrew Ng, free online materials on IR algorithms.
- Projects: Build simple ML pipelines and test retrieval techniques.

Step 4

Deep Dive into Natural Language Processing (NLP)



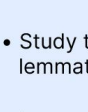
Hugging Face



ChatGPT



Claude



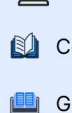
Actions

- Study text preprocessing: tokenization, stemming, lemmatization.
- Learn about embeddings: Word2Vec, Transformers.
- Understand and prompt large language models (LLMs): GPT-4, Claude 3.5, Gemini 1.5.
- Master prompt engineering for optimal query results.



Goals

- Gain expertise in text preprocessing and NLP models.



Resources

- Course: Introduction to NLP by Analytics Vidhya
- Guides: LangChain, LlamaIndex API documentation.
- Tools: NLTK, SpaCy, Hugging Face, ChatGPT

Step 5

Understand RAG Architecture



LangChain



LlamaIndex



Actions

- Understand the retriever-generator workflow.
- Learn how vector databases fetch relevant knowledge.
- Study how LLMs use retrieved-context for generating responses.



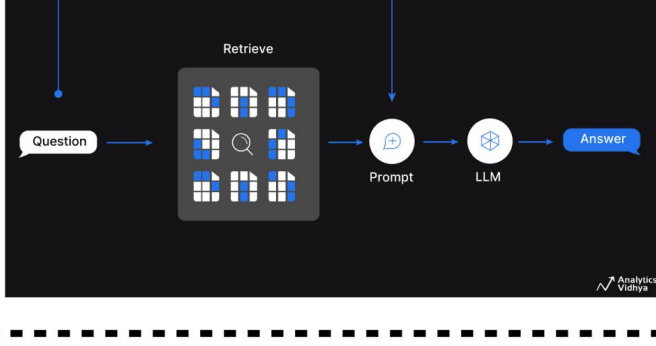
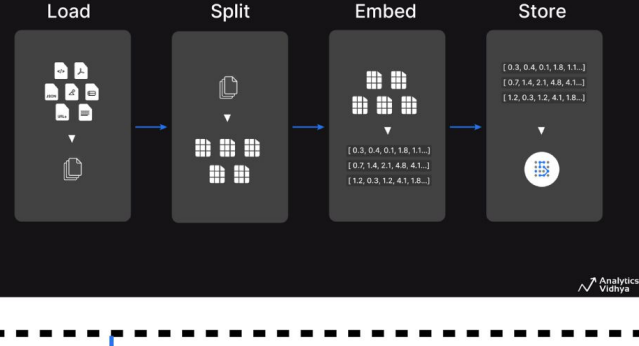
Goals

- Learn the two-stage architecture of RAG systems.



Resources

- Diagrams and whitepapers on RAG architecture.
- Practice creating RAG workflows using LangChain.
- Courses :Building Production-Ready RAG systems using LlamaIndex by Analytics Vidhya



Step 6

Implement and Evaluate RAG Systems



PHOENIX



LangSmith



ragas



Actions

- Start small: Build a question-answering system with RAG.
- Use tools like RAGAS and LangSmith for evaluation.
- Measure performance with metrics: contextual precision, recall, faithfulness.



Goals

- Gain hands-on experience in building and assessing RAG systems.

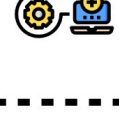
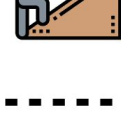


Resources

- Tools: LangSmith, Arize AI Phoenix, RAGAS, DeepEval
- Projects: Refine and optimize your RAG pipeline.
- Course: Building Your first RAG System using LlamaIndex - Free Course

Step 7

Address RAG Challenges



Actions

- Address hallucinations with better retrieval and reranking.
- Optimize embedding strategies for accuracy.
- Experiment with hybrid retrieval methods (dense + sparse approaches).
- Context compression strategies, like summarization or selective attention, filter out irrelevant information.
- Learn how to fine-tune hyperparameters for chunking and retrieval



Goals

- Learn to optimize RAG systems and overcome common challenges.

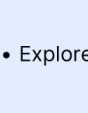


Resources

- Research papers on advanced retrieval strategies.
- Guides on fine-tuning LLMs.
- Course: Improving Real-World RAG Systems: Key Challenges & Practical Solutions by Analytics Vidhya

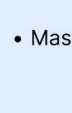
Step 8

Advance to Multimodal and Agentic RAG Systems



Actions

- Explore multimodal RAG systems (text, images, audio).
- Learn about agentic systems like Corrective RAG and Self-Reflective RAG.
- Develop systems for dynamic real-world applications (healthcare, legal, customer support).



Goals

- Master cutting-edge RAG advancements.



Resources

- Tutorials on multimodal integration.
- Case studies of agentic RAG implementations.
- Multimodal and Agentic RAG systems articles.

For More Amazing Infographics, Visit
www.analyticsvidhya.com