ATMA - algo(NEG-LLM) (1)

GETTING STARTED

USE CASES

h. GPT4-V Experiments with General, Specific questions and Chain Of Thought (COT) Prompting Technique.

i. Evaluating Multi-Modal RAG

j. Chroma Multi-Modal Demo with LlamaIndex

k. Multi-Modal on PDF’s with tables.

UNDERSTANDING

1. Building an LLM application
2. Using LLMs
3. Toggle child pages in navigation
    a. Privacy and Security
4. Loading Data (Ingestion)
5. Toggle child pages in navigation
    a. LlamaHub
    b. Documents / Nodes
    c. Toggle child pages in navigation
        i. Defining and Customizing Documents
        ii. Toggle child pages in navigation
            1. Metadata Extraction Usage Pattern
            2. Toggle child pages in navigation
                a. Extracting Metadata for Better Document Indexing and Understanding
                b. Automated Metadata Extraction for Better Retrieval + Synthesis
                c. Entity Metadata Extraction
                d. Metadata Extraction and Augmentation w/ Marvin
                e. Pydantic Extractor
        iii. Defining and Customizing Nodes

d. Node Parser Usage Pattern

e. Toggle child pages in navigation

i. Node Parser Modules

f. Ingestion Pipeline

g. Toggle child pages in navigation

i. Transformations

ii. Advanced Ingestion Pipeline

iii. Async Ingestion Pipeline + Metadata Extraction

iv. Ingestion Pipeline + Document Management

v. Redis Ingestion Pipeline

vi. Building a Live RAG Pipeline over Google Drive Files

6. Indexing
7. Storing
8. Querying
9. Putting It All Together
10. Toggle child pages in navigation

a. Q&A patterns

b. Toggle child pages in navigation

i. A Guide to Extracting Terms and Definitions

ii. A Guide to Creating a Unified Query Framework over your Indexes

iii. Knowledge Graphs

iv. Structured Data

v. A Guide to LlamaIndex + Structured Data

vi. Airbyte SQL Index Guide

c. Full-Stack Web Application

d. Toggle child pages in navigation

   a. Cost Analysis

   b. Toggle child pages in navigation

      i. Usage Pattern

OPTIMIZING

1. Basic Strategies

2. Toggle child pages in navigation

   a. Accessing/Customizing Prompts within Higher-Level Modules

   b. Advanced Prompt Techniques (Variable Mappings, Functions)

   c. Advanced Prompt Techniques (Variable Mappings, Functions)

   d. Prompt Engineering for RAG

   e. BM25 Retriever

   f. Reciprocal Rerank Fusion Retriever

   g. Weaviate Vector Store - Hybrid Search

   h. Pinecone Vector Store - Hybrid Search

   i. Vector Store Index usage examples

   j. Defining and Customizing Documents

   k. Toggle child pages in navigation

      i. Metadata Extraction Usage Pattern

      ii. Toggle child pages in navigation

         1. Extracting Metadata for Better Document Indexing and Understanding

         2. Automated Metadata Extraction for Better Retrieval + Synthesis

         3. Entity Metadata Extraction

         4. Metadata Extraction and Augmentation w/ Marvin

         5. Pydantic Extractor

   l. Defining and Customizing Nodes

- - m. Metadata Extraction Usage Pattern
  - n. Toggle child pages in navigation
    - i. Extracting Metadata for Better Document Indexing and Understanding
    - ii. Automated Metadata Extraction for Better Retrieval + Synthesis
    - iii. Entity Metadata Extraction
    - iv. Metadata Extraction and Augmentation w/ Marvin
    - v. Pydantic Extractor
- 3. Advanced Retrieval Strategies
- 4. Toggle child pages in navigation
  - a. Query Transform Cookbook
  - b. Query Transformations
  - c. Toggle child pages in navigation
    - i. HyDE Query Transform
    - ii. Multi-Step Query Engine
  - d. DeepMemory (Activeloop)
  - e. Weaviate Vector Store - Hybrid Search
  - f. Pinecone Vector Store - Hybrid Search
  - g. Structured Outputs
  - h. Toggle child pages in navigation
    - i. Pydantic Program
    - ii. Toggle child pages in navigation
      - 1. OpenAI Pydantic Program
      - 2. Guidance Pydantic Program
      - 3. Guidance for Sub-Question Query Engine
      - 4. DataFrame Structured Data Extraction
      - 5. Evaporate Demo

iii. Query Engines + Pydantic Outputs

iv. Toggle child pages in navigation

1. Query Engine with Pydantic Outputs
2. Pydantic Tree Summarize
3. Download Data

i. Pydantic Program

j. Toggle child pages in navigation

i. OpenAI Pydantic Program

ii. Guidance Pydantic Program

iii. Guidance for Sub-Question Query Engine

iv. DataFrame Structured Data Extraction

v. Evaporate Demo

k. Query Engines + Pydantic Outputs

l. Toggle child pages in navigation

i. Query Engine with Pydantic Outputs

ii. Pydantic Tree Summarize

iii. Download Data

5. Agentic strategies
6. Toggle child pages in navigation

a. Build your own OpenAI Agent

b. OpenAI Agent with Query Engine Tools

c. Retrieval-Augmented OpenAI Agent

d. OpenAI Agent + Query Engine Experimental Cookbook

e. OpenAI Agent Query Planning

f. Context-Augmented OpenAI Agent

7. Evaluation

8. Toggle child pages in navigation
   a. End-to-End Evaluation
   b. Toggle child pages in navigation
      i. QuestionGeneration
      ii. BatchEvalRunner - Running Multiple Evaluations
      iii. Correctness Evaluator
      iv. Faithfulness Evaluator
      v. Guideline Evaluator
      vi. Pairwise Evaluator
      vii. Relevancy Evaluator
      viii. Embedding Similarity Evaluator
   c. Component Wise Evaluation
   d. Toggle child pages in navigation
      i. BEIR Out of Domain Benchmark
      ii. HotpotQADistractor Demo
   e. Evaluating
   f. Toggle child pages in navigation
      i. Usage Pattern (Response Evaluation)
      ii. Usage Pattern (Retrieval)
      iii. Modules
      iv. Toggle child pages in navigation
         1. Faithfulness Evaluator
         2. Relevancy Evaluator
         3. Guideline Evaluator
         4. Correctness Evaluator
         5. Embedding Similarity Evaluator

9. Fine-tuning
10. Toggle child pages in navigation

    a. Fine-tuning an Adapter

    b. Embedding Fine-tuning Guide

    c. Router Fine-tuning

    d. Embedding Fine-tuning Repo

    e. Embedding Fine-tuning Blog

    f. GPT-3.5 Fine-tuning Notebook (Colab)

    g. GPT-3.5 Fine-tuning Notebook (Notebook link)

    h. Fine-tuning a gpt-3.5 ReAct Agent on Better Chain of Thought

    i. [WIP] Function Calling Fine-tuning

    j. GPT-3.5 Fine-tuning Notebook (Colab)

    k. GPT-3.5 Fine-tuning Notebook (in Repo)

    l. Fine-tuning with Retrieval Augmentation

    m. OpenAI Function Calling Fine-tuning

    n. Llama2 Structured Output Fine-tuning

    o. Fine-tuning to Memorize Knowledge

    p. Llama 2 Text-to-SQL Fine-tuning (w/ Gradient.AI)

    q. Llama 2 Text-to-SQL Fine-tuning (w/ Modal, Repo)

    r. Llama 2 Text-to-SQL Fine-tuning (w/ Modal, Notebook)

    s. Knowledge Distillation For Fine-Tuning A GPT-3.5 Judge (Correctness)

    t. Knowledge Distillation For Fine-Tuning A GPT-3.5 Judge (Pairwise)

    u. Cross-Encoder Finetuning

    v. Finetuning Llama 2 for Text-to-SQL

    w. Finetuning GPT-3.5 to Distill GPT-4

    x. Cohere Custom Reranker

11. Building Performant RAG Applications for Production
12. Toggle child pages in navigation
    a. Recursive Retriever + Query Engine Demo
    b. Document Summary Index
    c. Metadata Replacement + Node Sentence Window
    d. Auto-Retrieval from a Vector Database
    e. Document Summary Index
    f. Recursive Retriever + Document Agents
    g. Comparing Methods for Structured Retrieval (Auto-Retrieval vs. Recursive Retrieval)
    h. Sub Question Query Engine
    i. Joint QA Summary Query Engine
    j. Recursive Retriever + Document Agents
    k. Router Query Engine
    l. OpenAI Agent + Query Engine Experimental Cookbook
    m. OpenAI Agent Query Planning
    n. **Embedding Fine-tuning Guide**
13. Building RAG from Scratch (Lower-Level)
14. Toggle child pages in navigation
    a. Building Data Ingestion from Scratch
    b. Pinecone
    c. OpenAI
    d. Building Retrieval from Scratch
    e. Building RAG from Scratch (Open-source only!)
    f. Building a (Very Simple) Vector Store from Scratch
    g. Building Response Synthesis from Scratch

MODULE GUIDES

15. Llama API
16. LlamaCPP
17. LocalAI
18. MistralAI
19. Monster API LLM Integration into LLamaIndex
20. Ollama - Llama 2 7B
21. OpenAI
22. Azure OpenAI
23. OpenLLM
24. OpenRouter
25. PaLM
26. Perplexity
27. Portkey
28. Predibase
29. Replicate - Llama 2 13B
30. Replicate - Vicuna 13B
31. Llama2 + VectorStoreIndex
32. RunGPT
33. Setup
34. Getting Started
35. Basic auth example for service account
36. Streaming Usage
37. Chat Usage
38. Async Chat
39. Streaming Chat
40. Gemini Models

        1. Usage Pattern
        2. Completion prompts
        3. Chat prompts
        4. Accessing/Customizing Prompts within Higher-Level Modules
        5. Advanced Prompt Techniques (Variable Mappings, Functions)
        6. Prompt Engineering for RAG
        7. “Optimization by Prompting” for RAG
        8. EmotionPrompt in RAG

    ix. Using local models

    x. Run Llama2 locally

c. Embeddings

d. Toggle child pages in navigation

    i. OpenAI Embeddings

    ii. Langchain Embeddings

    iii. CohereAI Embeddings

    iv. Qdrant FastEmbed Embeddings

    v. Gradient Embeddings

    vi. Azure OpenAI

    vii. Custom Embeddings

    viii. Local Embeddings with HuggingFace

    ix. Elasticsearch Embeddings

    x. Embeddings with Clarifai

    xi. LLMRails Embeddings

    xii. Text Embedding Inference

    xiii. Google PaLM Embeddings

    xiv. Jina Embeddings

  - xv. Voyage Embeddings
  - xvi. MistralAI Embeddings
- e. [Beta] Multi-modal models
- f. Toggle child pages in navigation
  - i. Multi-Modal LLM using OpenAI GPT-4V model for image reasoning
  - ii. Multi-Modal LLM using Google’s Gemini model for image understanding and build Retrieval Augmented Generation with LlamaIndex
  - iii. Multi-Modal LLM using Replicate LlaVa, Fuyu 8B, MiniGPT4 models for image reasoning
  - iv. Multi-Modal GPT4V Pydantic Program
  - v. GPT4-V Experiments with General, Specific questions and Chain Of Thought (COT) Prompting Technique.
  - vi. Retrieval-Augmented Image Captioning
  - vii. Advanced Multi-Modal Retrieval using GPT4V and Multi-Modal Index/Retriever
  - viii. Multi-Modal on PDF’s with tables.
  - ix. Multi-Modal Retrieval using GPT text embedding and CLIP image embedding for Wikipedia Articles
  - x. Image to Image Retrieval using CLIP embedding and image correlation reasoning using GPT4V
  - xi. Chroma Multi-Modal Demo with LlamaIndex
  - xii. Evaluating Multi-Modal RAG
  - xiii. Retrieval Evaluation
- g. Evaluating With LabelledRagDataset’s
- h. Toggle child pages in navigation
  - i. Benchmarking RAG Pipelines With A LabelledRagDatatset
  - ii. Downloading a LlamaDataset from LlamaHub

i. Contributing A LabelledRagDataset

j. Toggle child pages in navigation

   i. LlamaDataset Submission Template Notebook

k. Component Wise Evaluation

l. Toggle child pages in navigation

   i. BEIR Out of Domain Benchmark

   ii. HotpotQADistractor Demo

m. End-to-End Evaluation

n. Toggle child pages in navigation

   i. QuestionGeneration

   ii. BatchEvalRunner - Running Multiple Evaluations

   iii. Correctness Evaluator

   iv. Faithfulness Evaluator

   v. Guideline Evaluator

   vi. Pairwise Evaluator

   vii. Relevancy Evaluator

   viii. Embedding Similarity Evaluator

o. Fine-tuning

p. Toggle child pages in navigation

   i. Fine-tuning an Adapter

   ii. Embedding Fine-tuning Guide

   iii. Router Fine-tuning

   iv. Embedding Fine-tuning Repo

   v. Embedding Fine-tuning Blog

   vi. GPT-3.5 Fine-tuning Notebook (Colab)

   vii. GPT-3.5 Fine-tuning Notebook (Notebook link)

viii. Fine-tuning a gpt-3.5 ReAct Agent on Better Chain of Thought

ix. [WIP] Function Calling Fine-tuning

x. GPT-3.5 Fine-tuning Notebook (Colab)

xi. GPT-3.5 Fine-tuning Notebook (in Repo)

xii. Fine-tuning with Retrieval Augmentation

xiii. OpenAI Function Calling Fine-tuning

xiv. Llama2 Structured Output Fine-tuning

xv. Fine-tuning to Memorize Knowledge

xvi. Llama 2 Text-to-SQL Fine-tuning (w/ Gradient.AI)

xvii. Llama 2 Text-to-SQL Fine-tuning (w/ Modal, Repo)

xviii. Llama 2 Text-to-SQL Fine-tuning (w/ Modal, Notebook)

xix. Knowledge Distillation For Fine-Tuning A GPT-3.5 Judge (Correctness)

xx. Knowledge Distillation For Fine-Tuning A GPT-3.5 Judge (Pairwise)

xxi. Cross-Encoder Finetuning

xxii. Finetuning Llama 2 for Text-to-SQL

xxiii. Finetuning GPT-3.5 to Distill GPT-4

xxiv. Cohere Custom Reranker

q. Building Performant RAG Applications for Production

r. Toggle child pages in navigation

i. Recursive Retriever + Query Engine Demo

ii. Document Summary Index

iii. Metadata Replacement + Node Sentence Window

iv. Auto-Retrieval from a Vector Database

v. Document Summary Index

vi. Recursive Retriever + Document Agents

vii. Comparing Methods for Structured Retrieval (Auto-Retrieval vs. Recursive Retrieval)

viii. Sub Question Query Engine

ix. Joint QA Summary Query Engine

x. Recursive Retriever + Document Agents

xi. Router Query Engine

xii. OpenAI Agent + Query Engine Experimental Cookbook

xiii. OpenAI Agent Query Planning

xiv. **Embedding Fine-tuning Guide**

s. Building RAG from Scratch (Lower-Level)

t. Toggle child pages in navigation

i. Building Data Ingestion from Scratch

ii. Pinecone

iii. OpenAI

iv. Building Retrieval from Scratch

v. Building RAG from Scratch (Open-source only!)

vi. Building a (Very Simple) Vector Store from Scratch

vii. Building Response Synthesis from Scratch

viii. Building Evaluation from Scratch

ix. Building a Router from Scratch

x. Building an Advanced Fusion Retriever from Scratch

MODULE GUIDES

1. Models
2. Toggle child pages in navigation

a. Using LLMs

b. Toggle child pages in navigation

i. Using LLMs as standalone modules

ii. Customizing LLMs within LlamaIndex Abstractions

iii. Available LLM integrations

iv. Toggle child pages in navigation

    1. AI21
    2. Anthropic
    3. Anyscale
    4. Bedrock
    5. Connect to Bedrock with Access Keys
    6. Clarifai LLM
    7. EverlyAI
    8. Gradient Base Model
    9. Gradient Model Adapter
    10. HuggingFace LLM - Camel-5b
    11. HuggingFace LLM - StableLM
    12. Local Llama2 + VectorStoreIndex
    13. LangChain LLM
    14. LiteLLM
    15. Llama API
    16. LlamaCPP
    17. LocalAI
    18. MistralAI
    19. Monster API LLM Integration into LLamaIndex
    20. Ollama - Llama 2 7B
    21. OpenAI
    22. Azure OpenAI

iii. Multi-Modal LLM using Replicate LlaVa, Fuyu 8B, MiniGPT4 models for image reasoning

iv. Multi-Modal GPT4V Pydantic Program

v. GPT4-V Experiments with General, Specific questions and Chain Of Thought (COT) Prompting Technique.

vi. Retrieval-Augmented Image Captioning

vii. Advanced Multi-Modal Retrieval using GPT4V and Multi-Modal Index/Retriever

viii. Multi-Modal on PDF’s with tables.

ix. Multi-Modal Retrieval using GPT text embedding and CLIP image embedding for Wikipedia Articles

x. Image to Image Retrieval using CLIP embedding and image correlation reasoning using GPT4V

xi. Chroma Multi-Modal Demo with LlamaIndex

xii. Evaluating Multi-Modal RAG

3. Prompts
4. Toggle child pages in navigation

   a. Usage Pattern

   b. Completion prompts

   c. Chat prompts

   d. Accessing/Customizing Prompts within Higher-Level Modules

   e. Advanced Prompt Techniques (Variable Mappings, Functions)

   f. Prompt Engineering for RAG

   g. “Optimization by Prompting” for RAG

   h. EmotionPrompt in RAG

5. Using local models
6. Run Llama2 locally

7. Embeddings
8. Toggle child pages in navigation
    a. OpenAI Embeddings
    b. Langchain Embeddings
    c. CohereAI Embeddings
    d. Qdrant FastEmbed Embeddings
    e. Gradient Embeddings
    f. Azure OpenAI
    g. Custom Embeddings
    h. Local Embeddings with HuggingFace
    i. Elasticsearch Embeddings
    j. Embeddings with Clarifai
    k. LLMRails Embeddings
    l. Text Embedding Inference
    m. Google PaLM Embeddings
    n. Jina Embeddings
    o. Voyage Embeddings
    p. MistralAI Embeddings
9. [Beta] Multi-modal models
10. Toggle child pages in navigation
    a. Multi-Modal LLM using OpenAI GPT-4V model for image reasoning
    b. Multi-Modal LLM using Google’s Gemini model for image understanding and build Retrieval Augmented Generation with LlamaIndex
    c. Multi-Modal LLM using Replicate LlaVa, Fuyu 8B, MiniGPT4 models for image reasoning
    d. Multi-Modal GPT4V Pydantic Program

    e. GPT4-V Experiments with General, Specific questions and Chain Of Thought (COT) Prompting Technique.
    f. Retrieval-Augmented Image Captioning
    g. Advanced Multi-Modal Retrieval using GPT4V and Multi-Modal Index/Retriever
    h. Multi-Modal on PDF’s with tables.
    i. Multi-Modal Retrieval using GPT text embedding and CLIP image embedding for Wikipedia Articles
    j. Image to Image Retrieval using CLIP embedding and image correlation reasoning using GPT4V
    k. Chroma Multi-Modal Demo with LlamaIndex
    l. Evaluating Multi-Modal RAG

11. Prompts
12. Toggle child pages in navigation
    a. Usage Pattern
    b. Completion prompts
    c. Chat prompts
    d. Accessing/Customizing Prompts within Higher-Level Modules
    e. Advanced Prompt Techniques (Variable Mappings, Functions)
    f. Prompt Engineering for RAG
    g. “Optimization by Prompting” for RAG
    h. EmotionPrompt in RAG
13. Loading Data
14. Toggle child pages in navigation
    a. Data Connectors (LlamaHub)
    b. Toggle child pages in navigation
        i. Usage Pattern

ii. Module Guides

iii. Toggle child pages in navigation

1. Simple Directory Reader
2. Psychic Reader
3. DeepLake Reader
4. Qdrant Reader
5. Discord Reader
6. MongoDB Reader
7. Chroma Reader
8. MyScale Reader
9. Faiss Reader
10. Obsidian Reader
11. Slack Reader
12. Web Page Reader
13. Pinecone Reader
14. Mbox Reader
15. MilvusReader
16. Notion Reader
17. Github Repo Reader
18. Google Docs Reader
19. Database Reader
20. Twitter Reader
21. Weaviate Reader
22. Make Reader
23. Deplot Reader Demo

c. Documents / Nodes

    d. Toggle child pages in navigation
        i. Defining and Customizing Documents
        ii. Toggle child pages in navigation
            1. Metadata Extraction Usage Pattern
            2. Toggle child pages in navigation
                a. Extracting Metadata for Better Document Indexing and Understanding
                b. Automated Metadata Extraction for Better Retrieval + Synthesis
                c. Entity Metadata Extraction
                d. Metadata Extraction and Augmentation w/ Marvin
                e. Pydantic Extractor
        iii. Defining and Customizing Nodes
    e. Node Parser Usage Pattern
    f. Toggle child pages in navigation
        i. Node Parser Modules
    g. Ingestion Pipeline
    h. Toggle child pages in navigation
        i. Transformations
        ii. Advanced Ingestion Pipeline
        iii. Async Ingestion Pipeline + Metadata Extraction
        iv. Ingestion Pipeline + Document Management
        v. Redis Ingestion Pipeline
        vi. Building a Live RAG Pipeline over Google Drive Files
15. Indexing
16. Toggle child pages in navigation
    a. Using VectorStoreIndex

b. Toggle child pages in navigation

    i. Metadata Extraction

    ii. Toggle child pages in navigation

        1. Extracting Metadata for Better Document Indexing and Understanding

        2. Automated Metadata Extraction for Better Retrieval + Synthesis

        3. Entity Metadata Extraction

        4. Metadata Extraction and Augmentation w/ Marvin

        5. Pydantic Extractor

    iii. Document Management

    iv. Vector Store Index usage examples

c. How Each Index Works

d. Module Guides

e. Toggle child pages in navigation

    i. VectorStoreIndex

    ii. Toggle child pages in navigation

        1. Metadata Extraction

        2. Toggle child pages in navigation

            a. Extracting Metadata for Better Document Indexing and Understanding

            b. Automated Metadata Extraction for Better Retrieval + Synthesis

            c. Entity Metadata Extraction

            d. Metadata Extraction and Augmentation w/ Marvin

            e. Pydantic Extractor

        3. Document Management

        4. Vector Store Index usage examples

        iii. Summary Index

        iv. Tree Index

        v. Keyword Table Index

        vi. Knowledge Graph Index

        vii. Custom Retriever combining KG Index and VectorStore Index

        viii. Knowledge Graph Query Engine

        ix. Knowledge Graph RAG Query Engine

        x. REBEL + Knowledge Graph Index

        xi. REBEL + Wikipedia Filtering

        xii. SQL Index

        xiii. SQL Query Engine with LlamaIndex + DuckDB

        xiv. Document Summary Index

        xv. The ObjectIndex Class

    f. Composability

    g. Toggle child pages in navigation

        i. Composable Graph Basic

        ii. Composable Graph with Weaviate

        iii. Composable Graph

17. Storing

18. Toggle child pages in navigation

    a. Customizing Storage

    b. Persisting & Loading Data

    c. Vector Stores

    d. Toggle child pages in navigation

        i. Astra DB

        ii. Simple Vector Store - Async Index Creation

iii. Azure CosmosDB MongoDB Vector Store

iv. Cassandra Vector Store

v. Chroma

vi. Azure Cognitive Search

vii. DashVector Vector Store

viii. DeepLake Vector Store

ix. DocArray Hnsw Vector Store

x. DocArray InMemory Vector Store

xi. Epsilla Vector Store

xii. LanceDB Vector Store

xiii. Metal Vector Store

xiv. Milvus Vector Store

xv. MyScale Vector Store

xvi. Elasticsearch Vector Store

xvii. Faiss Vector Store

xviii. MongoDB Atlas

xix. Neo4j vector store

xx. Opensearch Vector Store

xxi. Pinecone Vector Store

xxii. Pinecone Vector Store - Hybrid Search

xxiii. pgvecto.rs

xxiv. Postgres Vector Store

xxv. Redis Vector Store

xxvi. Qdrant Vector Store

xxvii. Qdrant Hybrid Search

xxviii. Rockset Vector Store

xxix. Simple Vector Store

xxx. Supabase Vector Store

xxxi. Tair Vector Store

xxxii. Tencent Cloud VectorDB

xxxiii. Timescale Vector Store (PostgreSQL)

xxxiv. Weaviate Vector Store

xxxv. Weaviate Vector Store - Hybrid Search

xxxvi. Zep Vector Store

e. Document Stores

f. Index Stores

g. Key-Value Stores

h. Using Graph Stores

i. Toggle child pages in navigation

i. Neo4j Graph Store

ii. Nebula Graph Store

iii. Knowledge Graph Query Engine

iv. Kuzu Graph Store

v. FalkorDB Graph Store

19. Querying

20. Toggle child pages in navigation

a. Query Engine

b. Toggle child pages in navigation

i. Usage Pattern

ii. Toggle child pages in navigation

1. Response Modes

2. Streaming

iii. Module Guides

iv. Toggle child pages in navigation

   1. Custom Query Engine
   2. Retriever Query Engine
   3. Text-to-SQL Guide (Query Engine + Retriever)
   4. JSON Query Engine
   5. Pandas Query Engine
   6. Knowledge Graph Query Engine
   7. Knowledge Graph RAG Query Engine
   8. Structured Hierarchical Retrieval
   9. Router Query Engine
   10. Retriever Router Query Engine
   11. Joint QA Summary Query Engine
   12. Sub Question Query Engine
   13. Multi-Step Query Engine
   14. SQL Router Query Engine
   15. SQL Auto Vector Query Engine
   16. SQL Join Query Engine
   17. [Beta] Text-to-SQL with PGVector
   18. SQL Query Engine with LlamaIndex + DuckDB
   19. Retry Query Engine
   20. CitationQueryEngine
   21. Recursive Retriever + Query Engine Demo
   22. Joint Tabular/Semantic QA over Tesla 10K
   23. Recursive Retriever + Document Agents
   24. Ensemble Query Engine Guide

        25. Sub Question Query Engine
        26. Recursive Retriever + Document Agents
        27. Multi-Document Agents
        28. Multi-Document Agents (V1)
        29. FLARE Query Engine
    v. Supporting Modules
    vi. Toggle child pages in navigation
        1. Query Transformations
        2. Toggle child pages in navigation
            a. HyDE Query Transform
            b. Multi-Step Query Engine
c. Chat Engine
d. Toggle child pages in navigation
    i. Usage Pattern
    ii. Module Guides
    iii. Toggle child pages in navigation
        1. ReAct Chat Engine
        2. OpenAI Chat Engine
        3. Condense Question Chat Engine
        4. Context Chat Engine
        5. Context Plus Condense Chat Engine
        6. Simple Chat Engine
e. Data Agents
f. Toggle child pages in navigation
    i. Usage Pattern
    ii. Module Guides

iii. Toggle child pages in navigation

   1. Build your own OpenAI Agent
   2. OpenAI Agent with Query Engine Tools
   3. Retrieval-Augmented OpenAI Agent
   4. OpenAI Agent + Query Engine Experimental Cookbook
   5. OpenAI Agent Query Planning
   6. Context-Augmented OpenAI Agent
   7. Recursive Retriever + Document Agents
   8. Multi-Document Agents
   9. GPT Builder Demo
   10. Single-Turn Multi-Function Calling OpenAI Agents
   11. OpenAI Assistant Agent
   12. Benchmarking OpenAI Retrieval API (through Assistant Agent)
   13. OpenAI Assistant Advanced Retrieval Cookbook
   14. ReAct Agent with Query Engine Tools
   15. Step-wise, Controllable Agents
   16. Controllable Agents for RAG
   17. Controllable Agents for RAG

iv. Tools

v. Toggle child pages in navigation

   1. Usage Pattern
   2. LlamaHub Tools Guide

vi. Lower-Level Agent API

g. Retriever

h. Toggle child pages in navigation

i. Retriever Modes

ii. Retriever Modules

iii. Toggle child pages in navigation

1. Define Custom Retriever
2. BM25 Hybrid Retriever
3. Simple Fusion Retriever
4. Reciprocal Rerank Fusion Retriever
5. Auto Merging Retriever
6. Metadata Replacement + Node Sentence Window
7. Auto Retriever (with Pinecone + Arize Phoenix)
8. Auto-Retrieval (with Chroma)
9. Auto-Retrieval (with BagelDB)
10. Structured Hierarchical Retrieval
11. Custom Retriever (KG Index and Vector Store Index)
12. Knowledge Graph RAG Retriever
13. Recursive Retriever + Query Engine Demo
14. Recursive Retriever + Node References
15. Recursive Retriever + Node References + Braintrust
16. Router Retriever
17. Ensemble Retrieval Guide
18. Google Generative Language Semantic Retriever
19. Structured Hierarchical Retrieval
20. Google Generative Language Semantic Retriever
21. Vectara Managed Index
22. Managed Index with Zilliz Cloud Pipeline
23. You.com Retriever
24. Text-to-SQL Guide (Query Engine + Retriever)

        25. DeepMemory (Activeloop)

i. Response Synthesizer

j. Toggle child pages in navigation

    i. Response Synthesis Modules

    ii. Toggle child pages in navigation

        1. Refine

        2. Refine with Structured Answer Filtering

        3. Tree Summarize

        4. Pydantic Tree Summarize

k. Routers

l. Toggle child pages in navigation

    i. Router Query Engine

    ii. Retriever Router Query Engine

    iii. SQL Router Query Engine

    iv. Router Retriever

m. Node Postprocessor

n. Toggle child pages in navigation

    i. Node Postprocessor Modules

    ii. Toggle child pages in navigation

        1. Sentence Embedding Optimizer

        2. Cohere Rerank

        3. LLM Reranker Demonstration (2021 Lyft 10-k)

        4. LLM Reranker Demonstration (Great Gatsby)

        5. Recency Filtering

        6. Time-Weighted Rerank

        7. PII Masking

    m. Google PaLM Embeddings

    n. Jina Embeddings

    o. Voyage Embeddings

    p. MistralAI Embeddings

25. [Beta] Multi-modal models

26. Toggle child pages in navigation

    a. Multi-Modal LLM using OpenAI GPT-4V model for image reasoning

    b. Multi-Modal LLM using Google’s Gemini model for image understanding and build Retrieval Augmented Generation with LlamaIndex

    c. Multi-Modal LLM using Replicate LlaVa, Fuyu 8B, MiniGPT4 models for image reasoning

    d. Multi-Modal GPT4V Pydantic Program

    e. GPT4-V Experiments with General, Specific questions and Chain Of Thought (COT) Prompting Technique.

    f. Retrieval-Augmented Image Captioning

    g. Advanced Multi-Modal Retrieval using GPT4V and Multi-Modal Index/Retriever

    h. Multi-Modal on PDF’s with tables.

    i. Multi-Modal Retrieval using GPT text embedding and CLIP image embedding for Wikipedia Articles

    j. Image to Image Retrieval using CLIP embedding and image correlation reasoning using GPT4V

    k. Chroma Multi-Modal Demo with LlamaIndex

    l. Evaluating Multi-Modal RAG

27. Prompts

28. Toggle child pages in navigation

    a. Usage Pattern

    b. Completion prompts
    c. Chat prompts
    d. Accessing/Customizing Prompts within Higher-Level Modules
    e. Advanced Prompt Techniques (Variable Mappings, Functions)
    f. Prompt Engineering for RAG
    g. “Optimization by Prompting” for RAG
    h. EmotionPrompt in RAG

29. Loading Data
30. Toggle child pages in navigation
    a. Data Connectors (LlamaHub)
    b. Toggle child pages in navigation
        i. Usage Pattern
        ii. Module Guides
        iii. Toggle child pages in navigation
            1. Simple Directory Reader
            2. Psychic Reader
            3. DeepLake Reader
            4. Qdrant Reader
            5. Discord Reader
            6. MongoDB Reader
            7. Chroma Reader
            8. MyScale Reader
            9. Faiss Reader
            10. Obsidian Reader
            11. Slack Reader
            12. Web Page Reader

      13. Pinecone Reader
      14. Mbox Reader
      15. MilvusReader
      16. Notion Reader
      17. Github Repo Reader
      18. Google Docs Reader
      19. Database Reader
      20. Twitter Reader
      21. Weaviate Reader
      22. Make Reader
      23. Deplot Reader Demo
c. Documents / Nodes
d. Toggle child pages in navigation
   i. Defining and Customizing Documents
   ii. Toggle child pages in navigation
      1. Metadata Extraction Usage Pattern
      2. Toggle child pages in navigation
         a. Extracting Metadata for Better Document Indexing and Understanding
         b. Automated Metadata Extraction for Better Retrieval + Synthesis
         c. Entity Metadata Extraction
         d. Metadata Extraction and Augmentation w/ Marvin
         e. Pydantic Extractor
   iii. Defining and Customizing Nodes
e. Node Parser Usage Pattern
f. Toggle child pages in navigation

    - i. Node Parser Modules
  - g. Ingestion Pipeline
  - h. Toggle child pages in navigation
    - i. Transformations
    - ii. Advanced Ingestion Pipeline
    - iii. Async Ingestion Pipeline + Metadata Extraction
    - iv. Ingestion Pipeline + Document Management
    - v. Redis Ingestion Pipeline
    - vi. Building a Live RAG Pipeline over Google Drive Files
- 31. Indexing
- 32. Toggle child pages in navigation
  - a. Using VectorStoreIndex
  - b. Toggle child pages in navigation
    - i. Metadata Extraction
    - ii. Toggle child pages in navigation
      - 1. Extracting Metadata for Better Document Indexing and Understanding
      - 2. Automated Metadata Extraction for Better Retrieval + Synthesis
      - 3. Entity Metadata Extraction
      - 4. Metadata Extraction and Augmentation w/ Marvin
      - 5. Pydantic Extractor
    - iii. Document Management
    - iv. Vector Store Index usage examples
  - c. How Each Index Works
  - d. Module Guides
  - e. Toggle child pages in navigation

        i. VectorStoreIndex

        ii. Toggle child pages in navigation

            1. Metadata Extraction

            2. Toggle child pages in navigation

                a. Extracting Metadata for Better Document Indexing and Understanding

                b. Automated Metadata Extraction for Better Retrieval + Synthesis

                c. Entity Metadata Extraction

                d. Metadata Extraction and Augmentation w/ Marvin

                e. Pydantic Extractor

            3. Document Management

            4. Vector Store Index usage examples

        iii. Summary Index

        iv. Tree Index

        v. Keyword Table Index

        vi. Knowledge Graph Index

        vii. Custom Retriever combining KG Index and VectorStore Index

        viii. Knowledge Graph Query Engine

        ix. Knowledge Graph RAG Query Engine

        x. REBEL + Knowledge Graph Index

        xi. REBEL + Wikipedia Filtering

        xii. SQL Index

        xiii. SQL Query Engine with LlamaIndex + DuckDB

        xiv. Document Summary Index

        xv. The ObjectIndex Class

    f. Composability

g. Toggle child pages in navigation

i. Composable Graph Basic

ii. Composable Graph with Weaviate

iii. Composable Graph

33. Storing

34. Toggle child pages in navigation

a. Customizing Storage

b. Persisting & Loading Data

c. Vector Stores

d. Toggle child pages in navigation

i. Astra DB

ii. Simple Vector Store - Async Index Creation

iii. Azure CosmosDB MongoDB Vector Store

iv. Cassandra Vector Store

v. Chroma

vi. Azure Cognitive Search

vii. DashVector Vector Store

viii. DeepLake Vector Store

ix. DocArray Hnsw Vector Store

x. DocArray InMemory Vector Store

xi. Epsilla Vector Store

xii. LanceDB Vector Store

xiii. Metal Vector Store

xiv. Milvus Vector Store

xv. MyScale Vector Store

xvi. Elasticsearch Vector Store

    xvii. Faiss Vector Store
    xviii. MongoDB Atlas
    xix. Neo4j vector store
    xx. Opensearch Vector Store
    xxi. Pinecone Vector Store
    xxii. Pinecone Vector Store - Hybrid Search
    xxiii. pgvecto.rs
    xxiv. Postgres Vector Store
    xxv. Redis Vector Store
    xxvi. Qdrant Vector Store
    xxvii. Qdrant Hybrid Search
    xxviii. Rockset Vector Store
    xxix. Simple Vector Store
    xxx. Supabase Vector Store
    xxxi. Tair Vector Store
    xxxii. Tencent Cloud VectorDB
    xxxiii. Timescale Vector Store (PostgreSQL)
    xxxiv. Weaviate Vector Store
    xxxv. Weaviate Vector Store - Hybrid Search
    xxxvi. Zep Vector Store

e. Document Stores

f. Index Stores

g. Key-Value Stores

h. Using Graph Stores

i. Toggle child pages in navigation

    i. Neo4j Graph Store

ii. Nebula Graph Store

iii. Knowledge Graph Query Engine

iv. Kuzu Graph Store

v. FalkorDB Graph Store

35. Querying
36. Toggle child pages in navigation
    - a. Query Engine
    - b. Toggle child pages in navigation
        - i. Usage Pattern
        - ii. Toggle child pages in navigation
            1. Response Modes
            2. Streaming
        - iii. Module Guides
        - iv. Toggle child pages in navigation
            1. Custom Query Engine
            2. Retriever Query Engine
            3. Text-to-SQL Guide (Query Engine + Retriever)
            4. JSON Query Engine
            5. Pandas Query Engine
            6. Knowledge Graph Query Engine
            7. Knowledge Graph RAG Query Engine
            8. Structured Hierarchical Retrieval
            9. Router Query Engine
            10. Retriever Router Query Engine
            11. Joint QA Summary Query Engine
            12. Sub Question Query Engine

        13. Multi-Step Query Engine
        14. SQL Router Query Engine
        15. SQL Auto Vector Query Engine
        16. SQL Join Query Engine
        17. [Beta] Text-to-SQL with PGVector
        18. SQL Query Engine with LlamaIndex + DuckDB
        19. Retry Query Engine
        20. CitationQueryEngine
        21. Recursive Retriever + Query Engine Demo
        22. Joint Tabular/Semantic QA over Tesla 10K
        23. Recursive Retriever + Document Agents
        24. Ensemble Query Engine Guide
        25. Sub Question Query Engine
        26. Recursive Retriever + Document Agents
        27. Multi-Document Agents
        28. Multi-Document Agents (V1)
        29. FLARE Query Engine
    v. Supporting Modules
    vi. Toggle child pages in navigation
        1. Query Transformations
        2. Toggle child pages in navigation
            a. HyDE Query Transform
            b. Multi-Step Query Engine
c. Chat Engine
d. Toggle child pages in navigation
    i. Usage Pattern

    ii. Module Guides
    iii. Toggle child pages in navigation
        1. ReAct Chat Engine
        2. OpenAI Chat Engine
        3. Condense Question Chat Engine
        4. Context Chat Engine
        5. Context Plus Condense Chat Engine
        6. Simple Chat Engine
e. Data Agents
f. Toggle child pages in navigation
    i. Usage Pattern
    ii. Module Guides
    iii. Toggle child pages in navigation
        1. Build your own OpenAI Agent
        2. OpenAI Agent with Query Engine Tools
        3. Retrieval-Augmented OpenAI Agent
        4. OpenAI Agent + Query Engine Experimental Cookbook
        5. OpenAI Agent Query Planning
        6. Context-Augmented OpenAI Agent
        7. Recursive Retriever + Document Agents
        8. Multi-Document Agents
        9. GPT Builder Demo
        10. Single-Turn Multi-Function Calling OpenAI Agents
        11. OpenAI Assistant Agent
        12. Benchmarking OpenAI Retrieval API (through Assistant Agent)
        13. OpenAI Assistant Advanced Retrieval Cookbook

14. ReAct Agent with Query Engine Tools

15. Step-wise, Controllable Agents

16. Controllable Agents for RAG

17. Controllable Agents for RAG

iv. Tools

v. Toggle child pages in navigation

1. Usage Pattern

2. LlamaHub Tools Guide

vi. Lower-Level Agent API

g. Retriever

h. Toggle child pages in navigation

i. Retriever Modes

ii. Retriever Modules

iii. Toggle child pages in navigation

1. Define Custom Retriever

2. BM25 Hybrid Retriever

3. Simple Fusion Retriever

4. Reciprocal Rerank Fusion Retriever

5. Auto Merging Retriever

6. Metadata Replacement + Node Sentence Window

7. Auto Retriever (with Pinecone + Arize Phoenix)

8. Auto-Retrieval (with Chroma)

9. Auto-Retrieval (with BagelDB)

10. Structured Hierarchical Retrieval

11. Custom Retriever (KG Index and Vector Store Index)

12. Knowledge Graph RAG Retriever

13. Recursive Retriever + Query Engine Demo
14. Recursive Retriever + Node References
15. Recursive Retriever + Node References + Braintrust
16. Router Retriever
17. Ensemble Retrieval Guide
18. Google Generative Language Semantic Retriever
19. Structured Hierarchical Retrieval
20. Google Generative Language Semantic Retriever
21. Vectara Managed Index
22. Managed Index with Zilliz Cloud Pipeline
23. You.com Retriever
24. Text-to-SQL Guide (Query Engine + Retriever)
25. DeepMemory (Activeloop)

i. Response Synthesizer

j. Toggle child pages in navigation

    i. Response Synthesis Modules

    ii. Toggle child pages in navigation

1. Refine
2. Refine with Structured Answer Filtering
3. Tree Summarize
4. Pydantic Tree Summarize

k. Routers

l. Toggle child pages in navigation

    i. Router Query Engine

    ii. Retriever Router Query Engine

    iii. SQL Router Query Engine

- iv. Router Retriever

m. Node Postprocessor

n. Toggle child pages in navigation

- i. Node Postprocessor Modules
- ii. Toggle child pages in navigation
    1. Sentence Embedding Optimizer
    2. Cohere Rerank
    3. LLM Reranker Demonstration (2021 Lyft 10-k)
    4. LLM Reranker Demonstration (Great Gatsby)
    5. Recency Filtering
    6. Time-Weighted Rerank
    7. PII Masking
    8. Forward/Backward Augmentation
    9. Metadata Replacement + Node Sentence Window
    10. LongContextReorder

o. Output Parsing Modules

p. Toggle child pages in navigation

- i. Guardrails Output Parsing
- ii. Langchain Output Parsing
- iii. Guidance Pydantic Program
- iv. Guidance for Sub-Question Query Engine
- v. OpenAI Pydantic Program