Agenda

① Ollama $\longrightarrow$ LLM's Local $\Rightarrow$ Open Source Models.

② Document reader, Text Splitter, Vector Databases, Embedding

OpenAI Ollama Hugginface

Open Source LLM's

Webbased loader

| Website |

Embeddings
[0.4 0.2 . . . . .]

[   ] $\Rightarrow$ | FAISS |

[   ]

| query | $\longrightarrow$ | LLM | $\longrightarrow$ O/p.

Prompt Template.

O/p Tokens

Context size of LLM's

| ChatOpenAI |

| CHATGROQ |

| CHATOLLAMA |

[
Answer the question

{ Context { ≡ } }

}
]

$\Rightarrow$ Documentchain
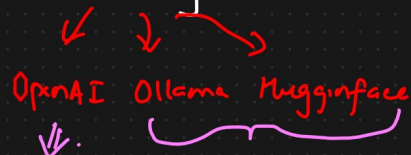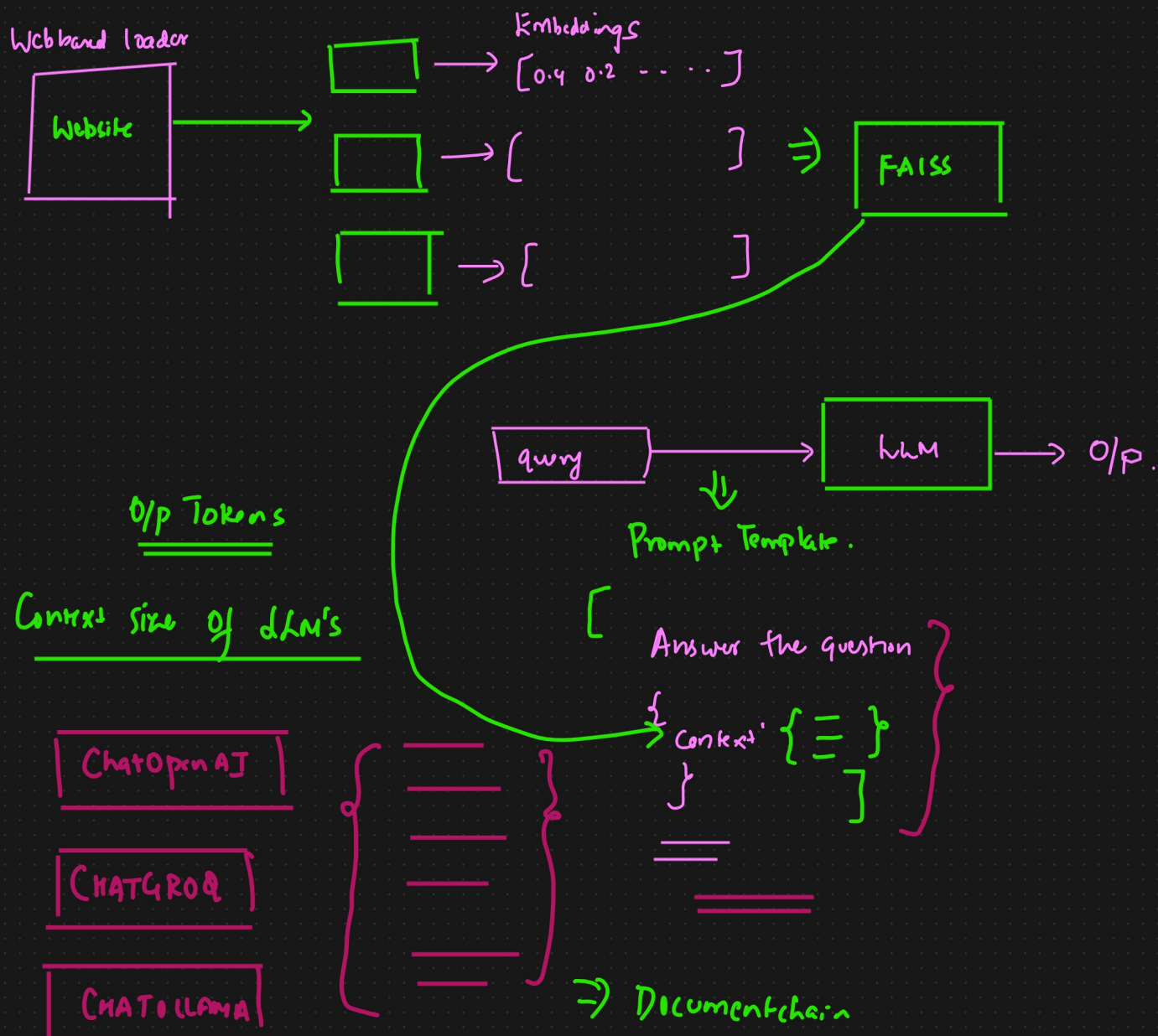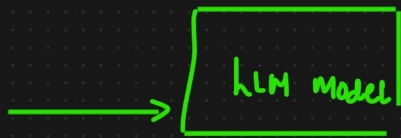
```
┌─────────────┐
──────────────→│  LLM model  │                    Exam
               └─────────────┘                                    ✓
                                         What is  Agentic AI?        8th chapter
   {
       ↗ ↑↑↑↑                     Ans)    1-4th chapters      ┌──────────────┐
   { Context }⇐                ⇒   O/p            ───────      │  8 chapters  │
           }                       ───           1-8th chapters └──────────────┘
           }                                        ↓↓
                                                              3rd Gen AI


                  ← Similarity Search           →Document stuff chain
   Vectrstore          ─────────                 ──────────────────
      ⇓                   ⇓
   {                   Indepent
   { RunnableBinding   Event      Runnable    | Prompt | llm | Stroutputparser
      ⇓                           Assign           ↗  ↗ ↗    ↗    ↗
   Retriever  → Interface                         ↗  ↗  ↗   ↗   ↗
      ══
   [ Vectrstore · as_retriever ]⇐                    ───→    ───→ O/p.
      ⇓

   Document  list  →①  }
                    ②  }
                    ③  }
                    ④  }
```
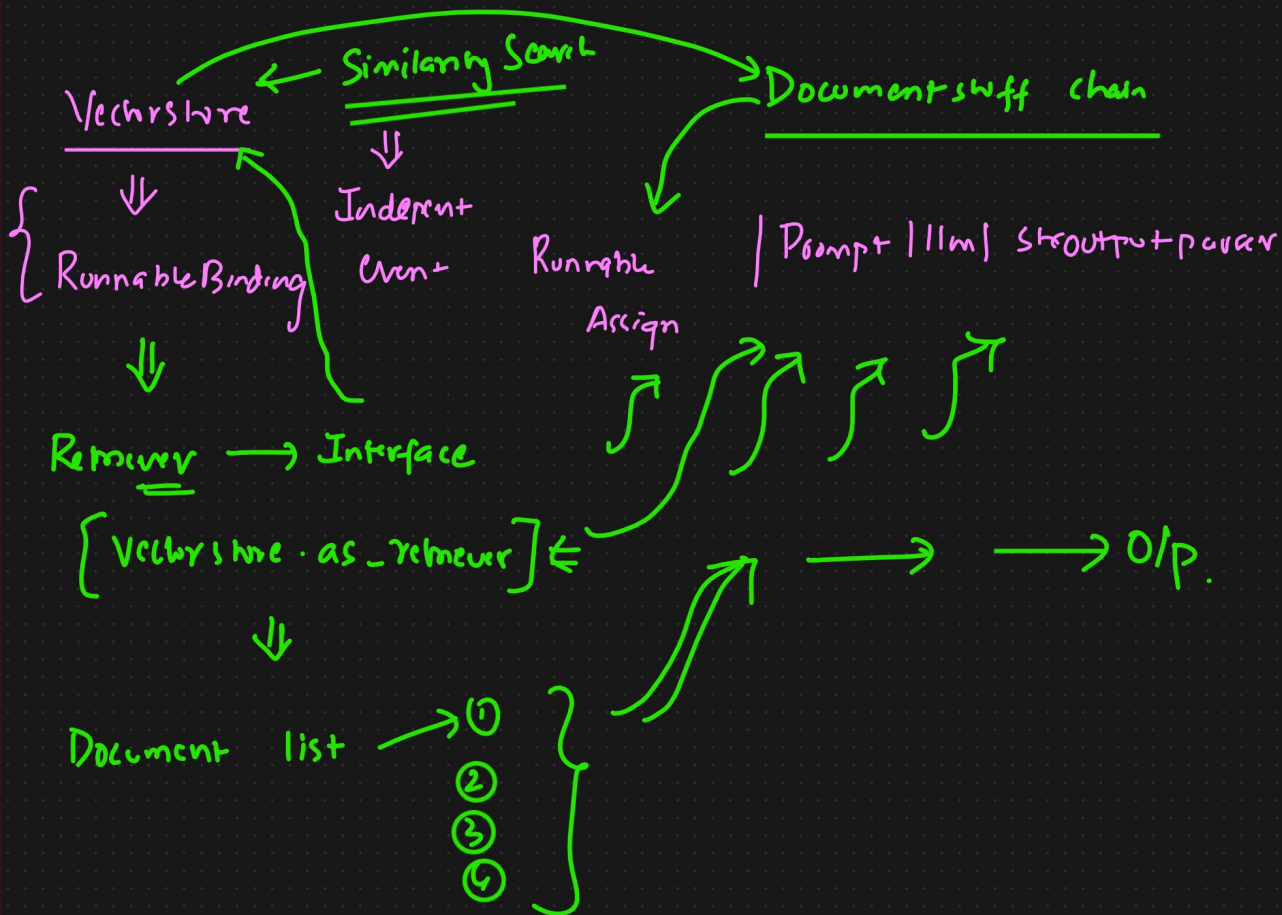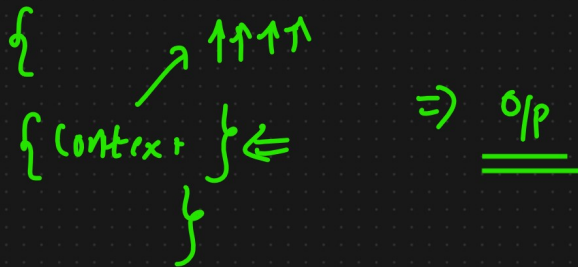
Basic RAG

Data Source

Step 1

DATA → Chunks → Embedding → [ ] ← Vectorstore

Custom

I/p → LLM

Chat Prompt

{ ——→ Prompt

{ Context }

}

retriever chain

Retriever → Stuff Document chain | Prompt | LLM | StrOutput()

List of Documents }

Runnable binding

{ Chain }