=> Chat GPT => Development
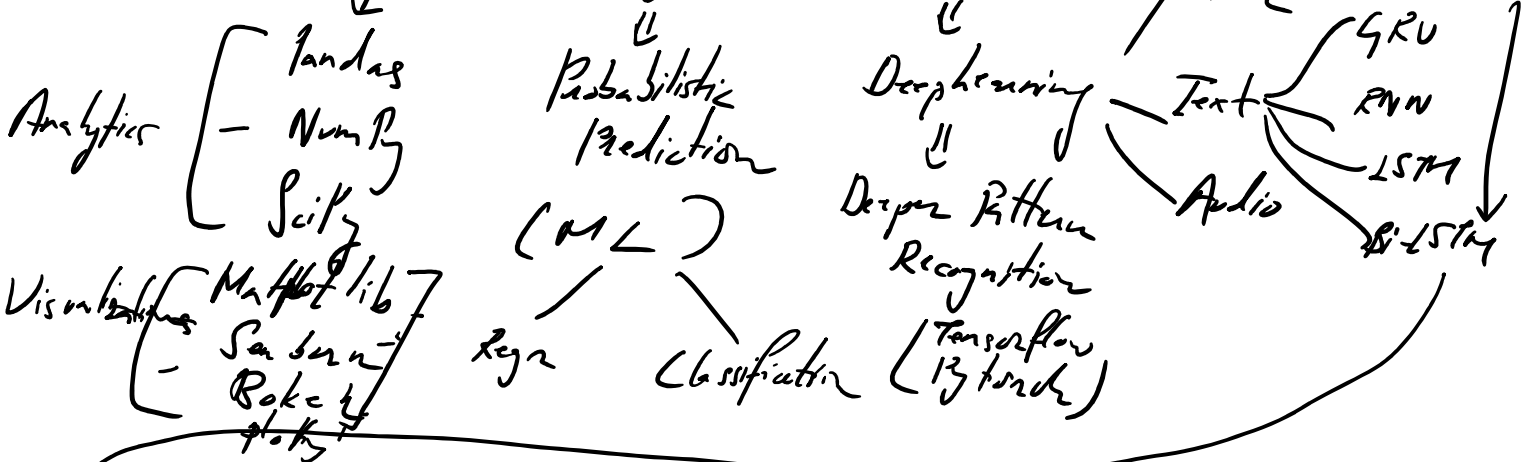
Facebook (Meta), Anthropic, Mixtral

Llama — Claude Sonnet

Large Language Models

=> Data => Information => (Knowledge) <= Condensed form

LLM

=> Corpus ~ Large Body of Text

=> Python => Analytics => Predictive Analytics => Non-parametric Approach

Analytics [Pandas, NumPy, SciPy]

Predictive Analytics => Probabilistic Prediction

Non-parametric Approach => Deep Learning => Deeper Pattern Recognition

Vision / Text / Audio

GRU, RNN, LSTM, BiLSTM

Visualization [Matplotlib, Seaborn, Bokeh, plotly]

(ML)

Regr   Classification   [Tensorflow, Pytorch]

Training Speeds
Accuracy
Effectiveness
} Transformer (2019)

(Attention is all you need)   (Bleu)

BERT
RoBERTa (Meta)
DistillNet

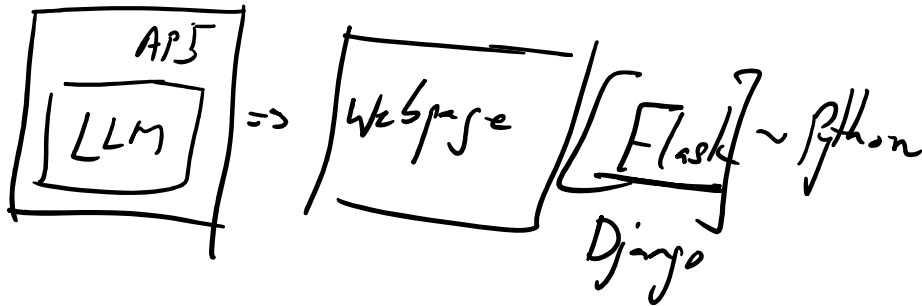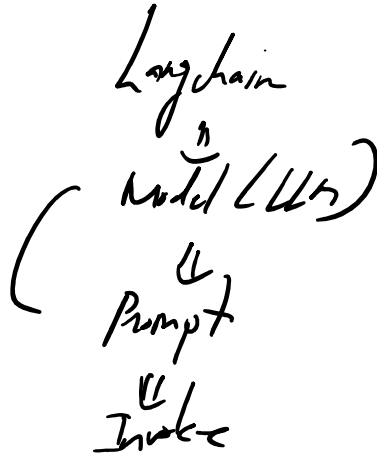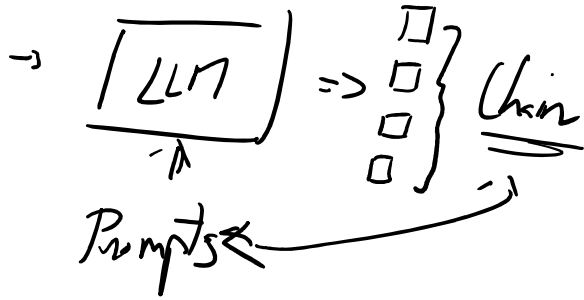=> Open AI (Wikipedia) 6 Billion Tokens => Large Language Model

(6 0 0 0 0 0 0 0 0 0)

$[0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$

Chat ~~GPT~~ GPT 1.0

→ Ollama ~ Interactively

⇒ Prompt Engineering

→ [LLM] ⇒ □ □ □ □ } Chain

Prompts

Langchain
↑
Model (LLM)
↓
Prompt
↓
Invoke

API
[LLM] ⇒ Webpage | [Flask] ~ Python
Django

FastApi

⇒ CRUD
(Get, Post, Put, Delete]

REST   API

[Server] ⟨⟨===⟩⟩ [Webpage]

Business
Logic

□ [Mobile Apps]

(Js)
Server → (JS) HTML/CSS
[::::]
PHP / JS / Ruby / Python / Java   Angular

Server => [diagram]
(Java)

Server => Android App
(logic)

Server => iOS App

PHP/JS/Ruby/Python/Java → Angular
Server [JSON] → [diagram] React
  < → > Kotlin/Java
JavaScript Object Notation iOS
(React Native & Flutter)
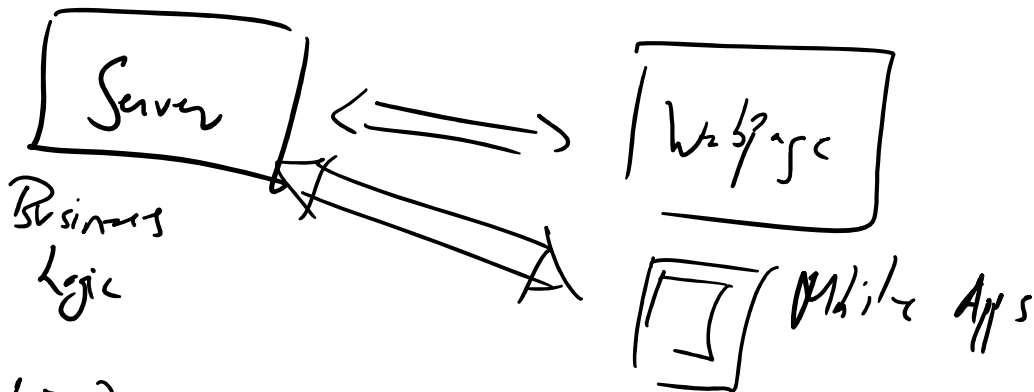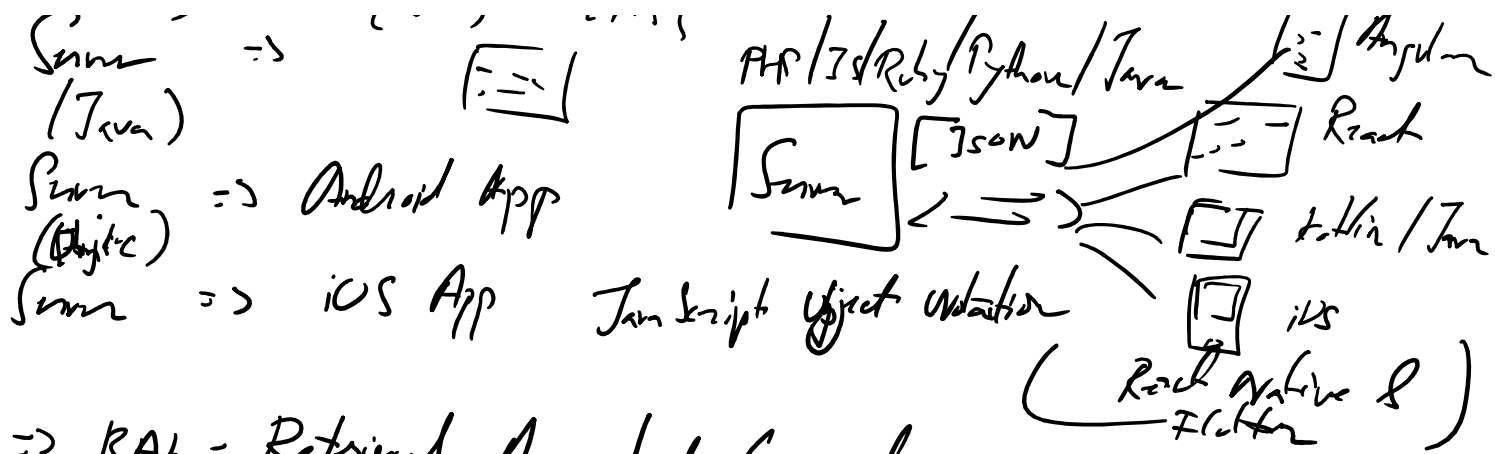
=> RAG - Retrieval Augmented Generation

→ Retrieval Chain

→ Vectorstores (Database)
    FAISS, ObjectBox, Pinecone

→ Language
    Vocab
    Grammar   ~ Embeddings
    Sentention   (Word2Vec, Glove)
    Accent

RAG Based System:

[Data Ingestion] — Manage Embeddings → [Vector Database] → [LLM] GenAI

Research Paper

Embeddings

Data ⟹ Information → Knowledge