INTERIM

Synapse

WHERE AI MEETS EDUCATION

PRESENTATION

Dr. Mary Priya Sebastian

Gautham C Sudheer
Fathima Jennath N K
Godwin Gino
Mohammed Basil

CONTENT

Problem Definition

The traditional process of converting physical text into digital, editable, and searchable formats is often time-consuming and inefficient, particularly for students and educators working with handwritten notes or multilingual content.

Purpose and Need

The project aims to meet the growing demand for digitized, interactive educational content by leveraging AI-powered tools.

- **Enhance Accessibility**: Converting text images, and multilingual content, into editable and searchable formats
- **Streamline Learning and Note-Taking**: Digitize and organize notes, to interact with content through features like AI-powered chatbots.
- **Promote Inclusive and Interactive Education**: With multilingual support and speech-to-text/text-to-speech functionalities

Objectives

01

Develop an AI-powered Platform For Text Digitization

Efficiently converts text images into editable and searchable digital content using advanced OCR technology.

02

Integrate Natural Language Processing (NLP) Capabilities

Enable interactive features, such as an AI chatbot for answering questions, generating summaries, and assisting with note organization.

03

Provide Multilingual Support and Accessibility Features

Speech-to-Text and Text-to-Speech, ensuring the platform is adaptable to users from diverse linguistic backgrounds and varying learning needs.

Literature Survey

Paper	Advantages	Disadvantages
S. Lei and Y. Li, "English Machine Translation System Based on Neural Network Algorithm," Procedia Computer Science, vol. 228, pp. 409-420, 2023	Better handling of long-range dependencies, parallel processing for faster training	High computational and memory requirements, for long sequences, and they can be data-hungry
A. T. Neumann, Y. Yin, S. Sowe, S. Decker, and M. Jarke, "An LLM-driven chatbot in higher education for databases and information systems," IEEE Transactions on Education, vol. 1, pp. 1-15, 2024	Provides personalized, quick responses to help students understand course material	The chatbot sometimes gives incorrect or repetitive answers, meaning it still needs better fact-checking to ensure accuracy
N. Sarika, N. Sirisala, and M. S. Velpuru, “Cnn based optical character recognition and applications,” Proc. Sixth Int. Conf. Inventive Comput. Technol. (ICICT 2021), pp. 666-672, 2021	VGG-16 model has shown 92% accuracy in Telugu handwritten character recognition	It requires more computational resources and longer training times, making it less efficient for scenarios with limited computational power.

Literature Survey

Paper	Advantages	Disadvantages
Rayyan Najam and Safiullah Faizullah, “Analysis of Recent Deep Learning Techniques for Arabic Handwritten-Text OCR and Post-OCR Correction” Applied Sciences, vol. 13, no. 13, p. 7568, Jun. 2023	Advanced architectures like Transformer-based models and RNNs can capture contextual relationships between characters, words, and lines	Deep learning models require large labeled datasets of handwritten text to achieve high accuracy
Lorenz Kuhn, Yarin Gal, Sebastian Farquhar, “Semantic Uncertainty: Linguistic Invariances for Uncertainty Estimation in Natural Language Generation” ICLR 2023, https://doi.org/10.48550/arXiv.2302.09664	Semantic Entropy captures uncertainty over meanings, not just forms, which provides a more reliable measure in tasks like question answering	The bidirectional entailment algorithm for clustering sentences has quadratic complexity; The performance is highly sensitive to sampling temperature and methods

Proposed Method

Text Recognition

Extracts and digitizes textbook and handwritten notes for easy editing and searching

Text Extraction

Converts images of text into editable and searchable digital content using advanced OCR technology

Note Organization

Categorizes and organizes digital notes based on content and user-defined criteria.

Multilingual Support

Translates text into multiple languages to cater to diverse linguistic backgrounds.

Proposed Method

Interactive Chatbot

Provides conversational interaction with notes, including answering questions and generating summaries

Text Summarization

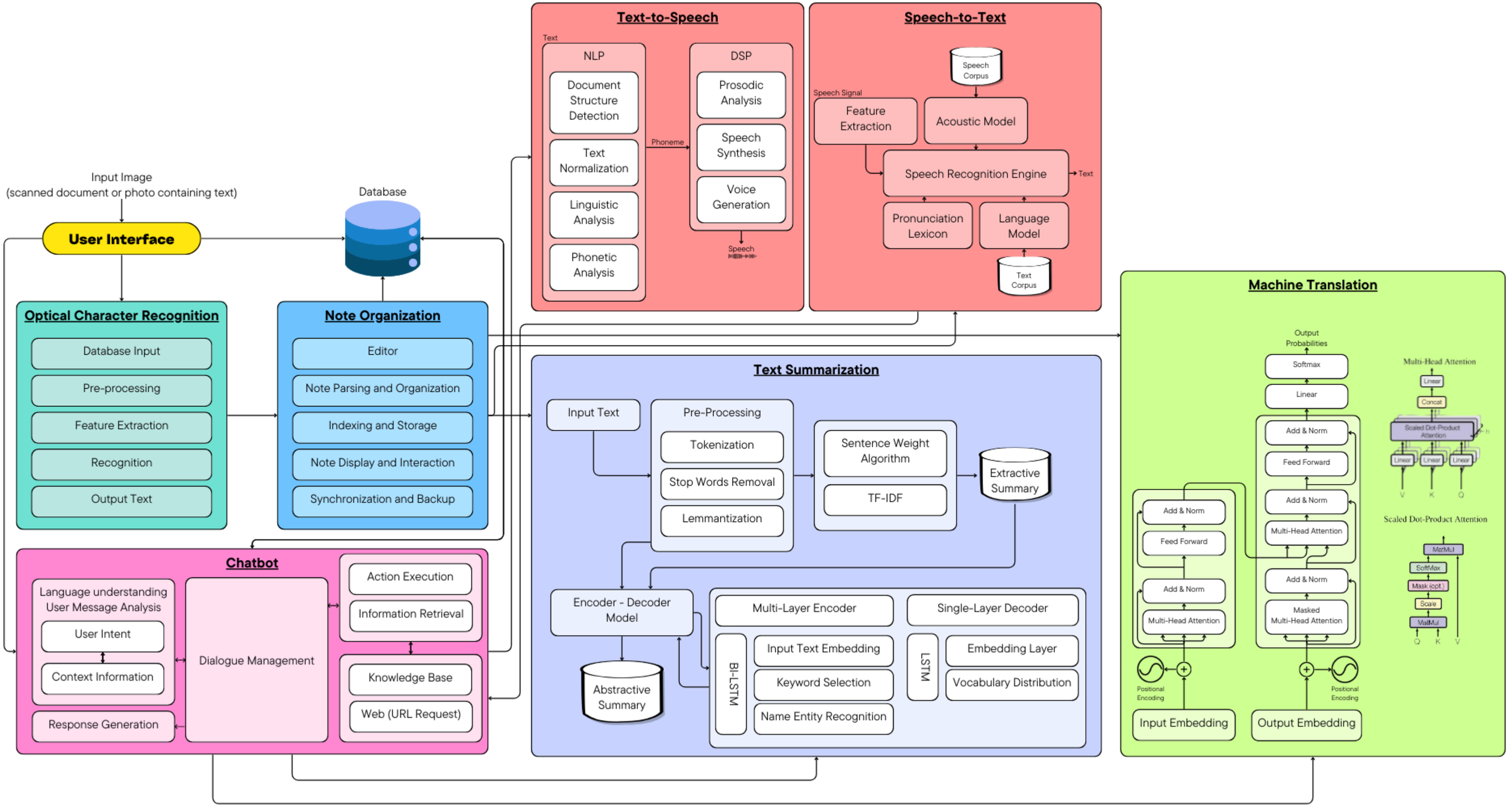
Automatically summarizes lengthy texts to highlight key points and essential information.

Speech-to-Text

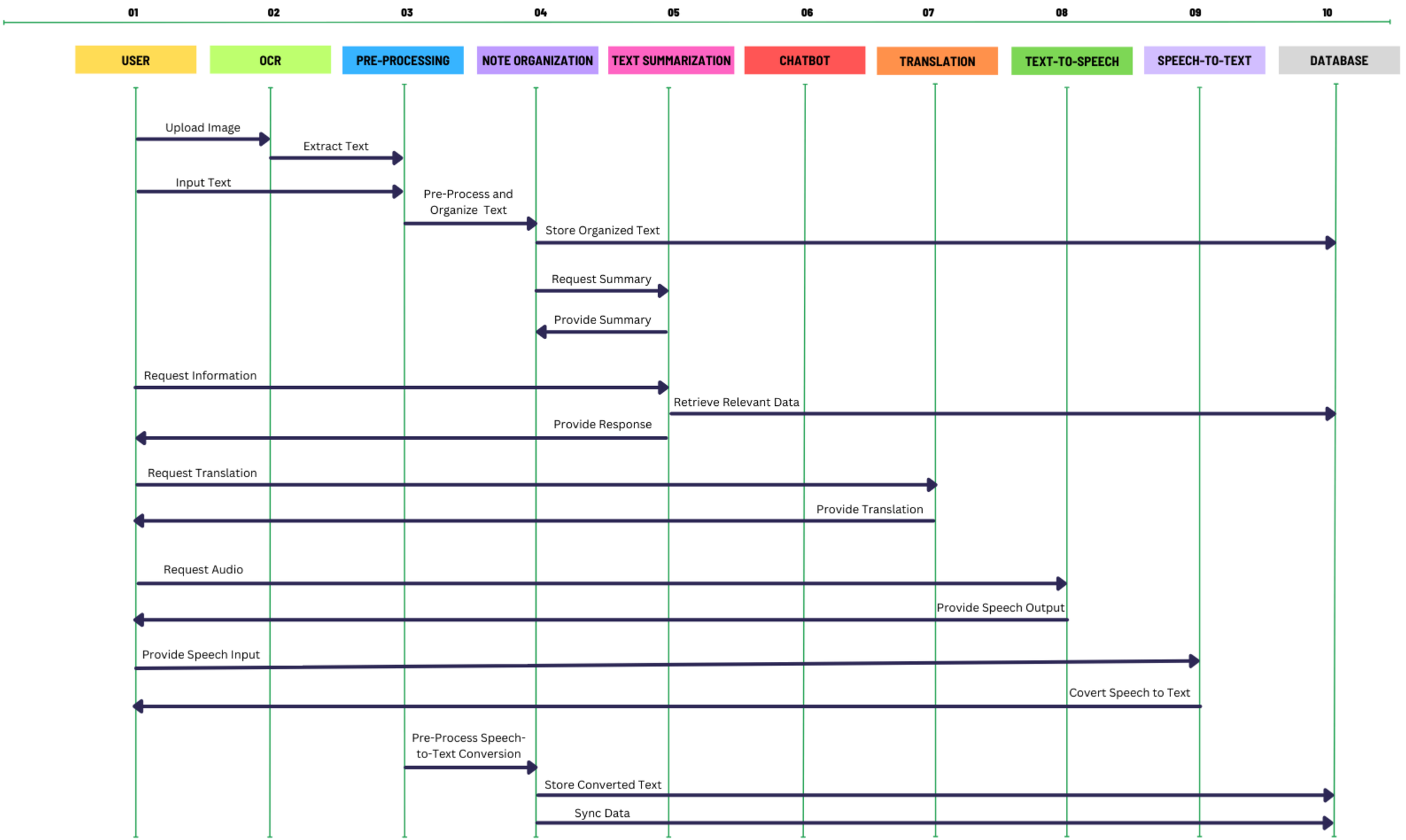
Converts spoken language into written text for improved accessibility and interactive learning

Text-to-Speech

Reads aloud text content to support auditory learning and accessibility needs.


Input Image
(scanned document or photo containing text)
User Interface
Database
Optical Character Recognition
Database Input
Pre-processing
Feature Extraction
Recognition
Output Text
Note Organization
Editor
Note Parsing and Organization
Indexing and Storage
Note Display and Interaction
Synchronization and Backup
Chatbot
Language understanding
User Message Analysis
User Intent
Context Information
Response Generation
Dialogue Management
Action Execution
Information Retrieval
Knowledge Base
Web (URL Request)
Text-to-Speech
Text
NLP
Document Structure Detection
Text Normalization
Linguistic Analysis
Phonetic Analysis
Phoneme
DSP
Prosodic Analysis
Speech Synthesis
Voice Generation
Speech
Speech-to-Text
Speech Corpus
Speech Signal
Feature Extraction
Acoustic Model
Speech Recognition Engine
Text
Pronunciation Lexicon
Language Model
Text Corpus
Text Summarization
Input Text
Pre-Processing
Tokenization
Stop Words Removal
Lemmantization
Sentence Weight Algorithm
TF-IDF
Extractive Summary
Encoder - Decoder Model
Abstractive Summary
Multi-Layer Encoder
Single-Layer Decoder
BI-LSTM
Input Text Embedding
Keyword Selection
Name Entity Recognition
LSTM
Embedding Layer
Vocabulary Distribution
Machine Translation
Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Positional Encoding
Input Embedding
Output Embedding
Multi-Head Attention
Concat
Scaled Dot-Product Attention
Linear
Linear
Linear
V
K
Q
Scaled Dot-Product Attention
MatMul
SoftMax
Mask (opt.)
Scale
MatMul
Q
K
V

SEQUENCE DIAGRAM



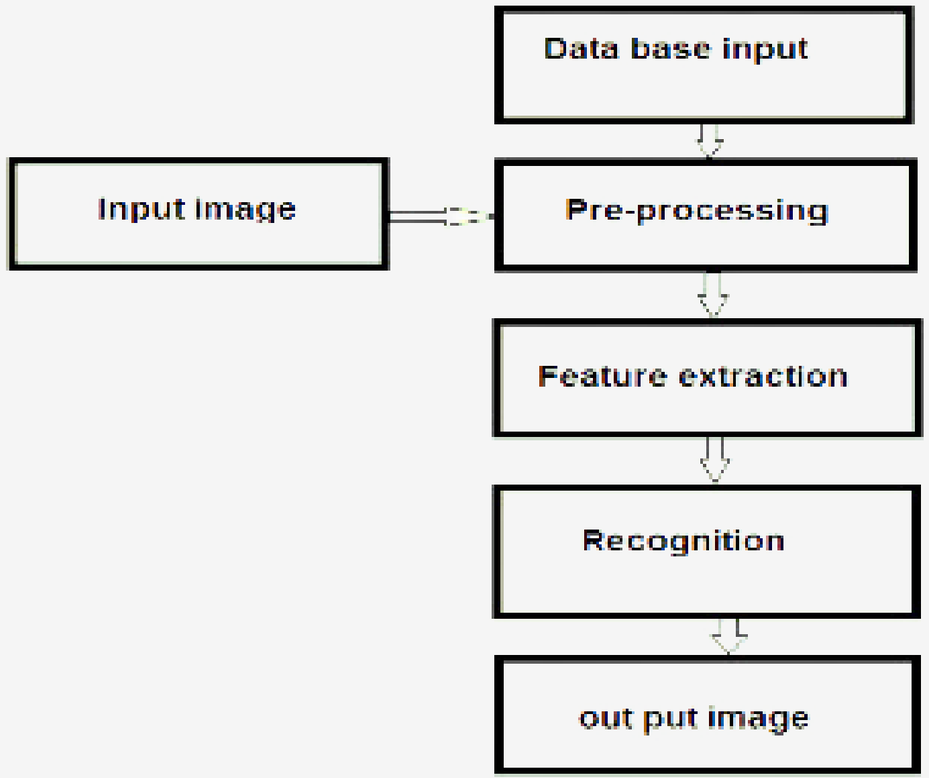
MODULE WISE DIAGRAMS

1. OCR & Text Extraction

- Grayscale Conversion, Binarization using Otsu’s thresholding, Noise Reduction using Gaussian blur or median filters, Skew Correction using Hough Transform
- Deep Learning models like EAST (Efficient and Accurate Scene Text Detector) to locate regions in the image containing text
- A Convolutional Recurrent Neural Network (CRNN) architecture combining CNNs for feature extraction and Long Short-Term Memory (LSTM) networks for sequence modeling.

1. OCR & Text Extraction

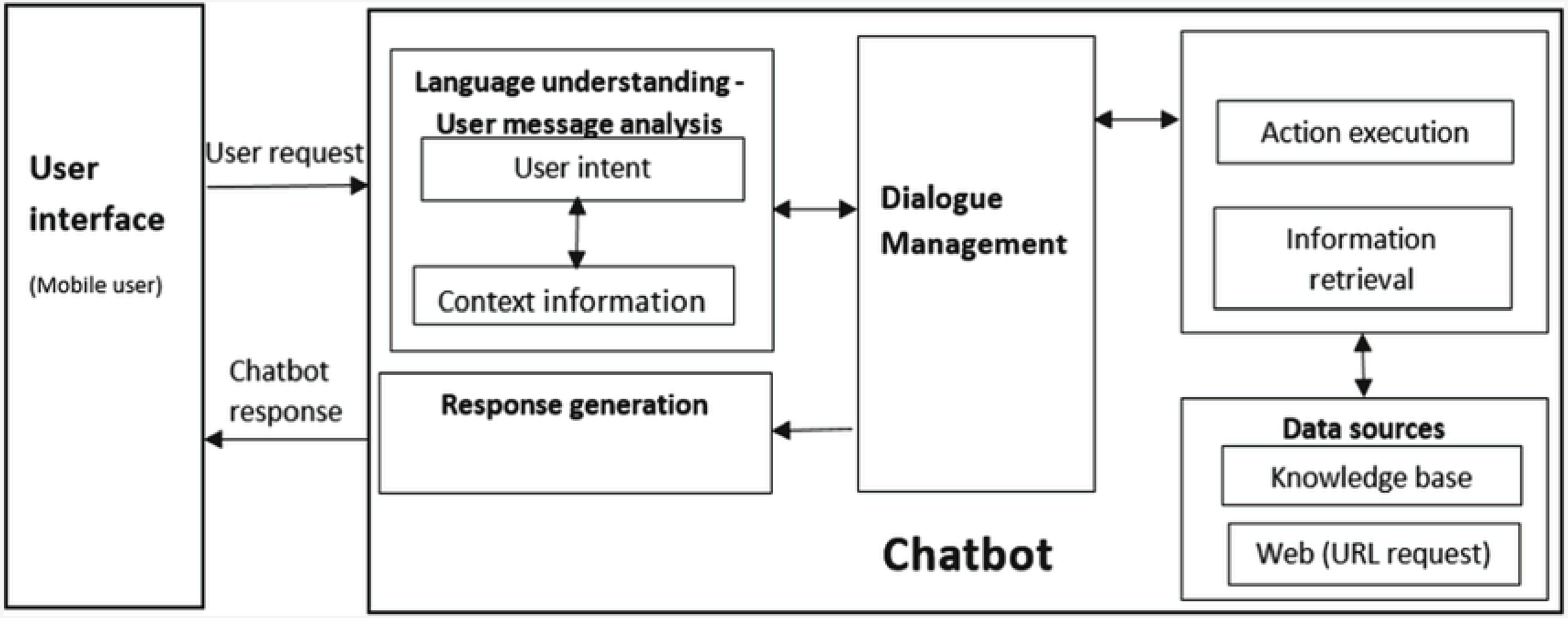
- Build a microservice that handles image uploads, processes them with OCR, and returns editable text.
- Image processing techniques and deep learning models to recognize and extract text accurately



2. AI-Powered Chatbot

- Natural Language Understanding: Tokenization and Named Entity Recognition (NER) using NLP libraries like SpaCy or Hugging Face's transformers.
- BERT or DistilBERT for context-aware understanding of user queries.
- A classification model (e.g., a fine-tuned BERT or RNN-based model) is used to determine the user's intent
- Entities and context are extracted to provide precise responses.
- The chatbot uses GPT or Dialogflow to generate context-aware responses.

2. AI-Powered Chatbot



3. Text Summarization

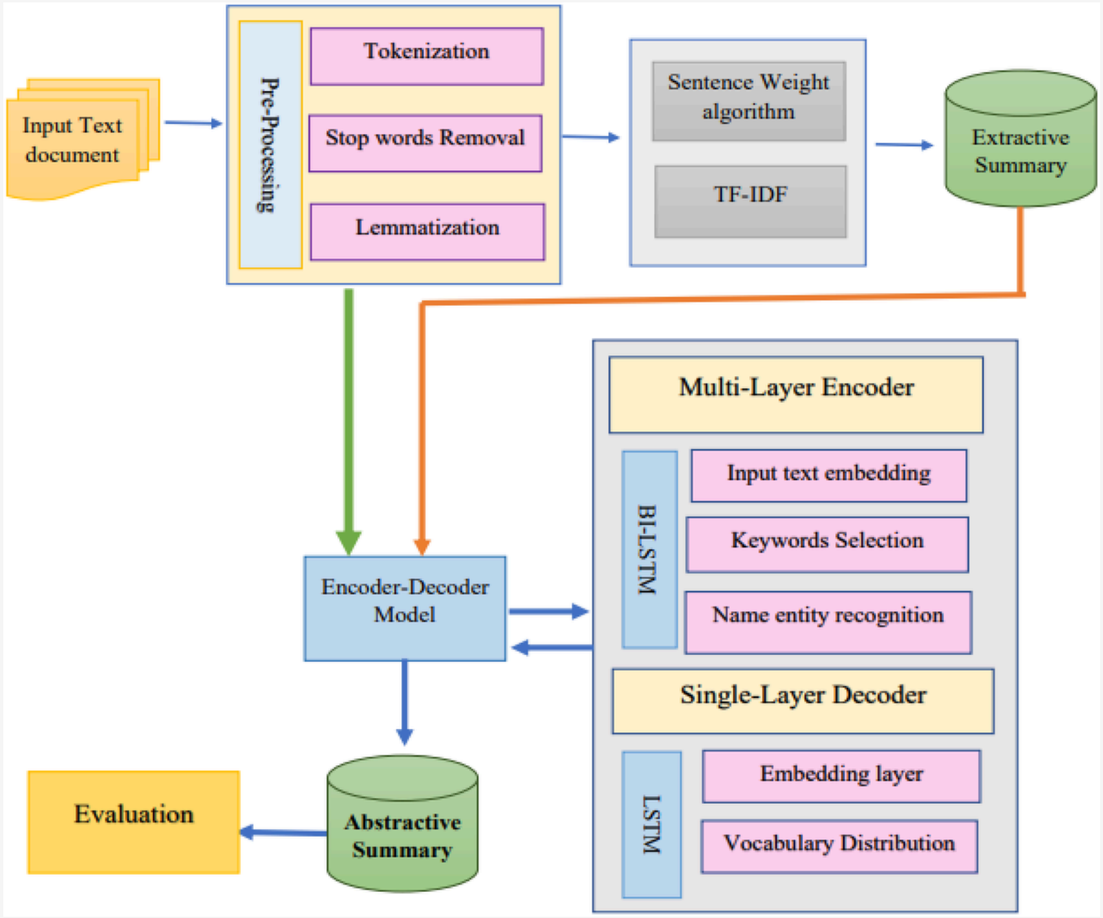
- Tokenization breaks the text into individual words and sentences
- Lemmatization & Stopword Removal simplifies words to their base forms and removes irrelevant words to focus on meaningful content
- Pre-trained word embeddings (e.g., Word2Vec, GloVe) or contextual embeddings from models like BERT (Bidirectional Encoder Representations from Transformers) to convert text into numerical vectors

3. Text Summarization

- Extractive Summarization: Algorithms like TextRank (a graph-based approach similar to PageRank) or deep learning models like BERT-based extractors to identify key sentences and rank them based on importance
- Abstractive Summarization: Transformer-based models like GPT (Generative Pre-trained Transformer) or BART (Bidirectional and Auto-Regressive Transformers) to generate new, shorter text summaries that maintain the original meaning.

3. Text Summarization

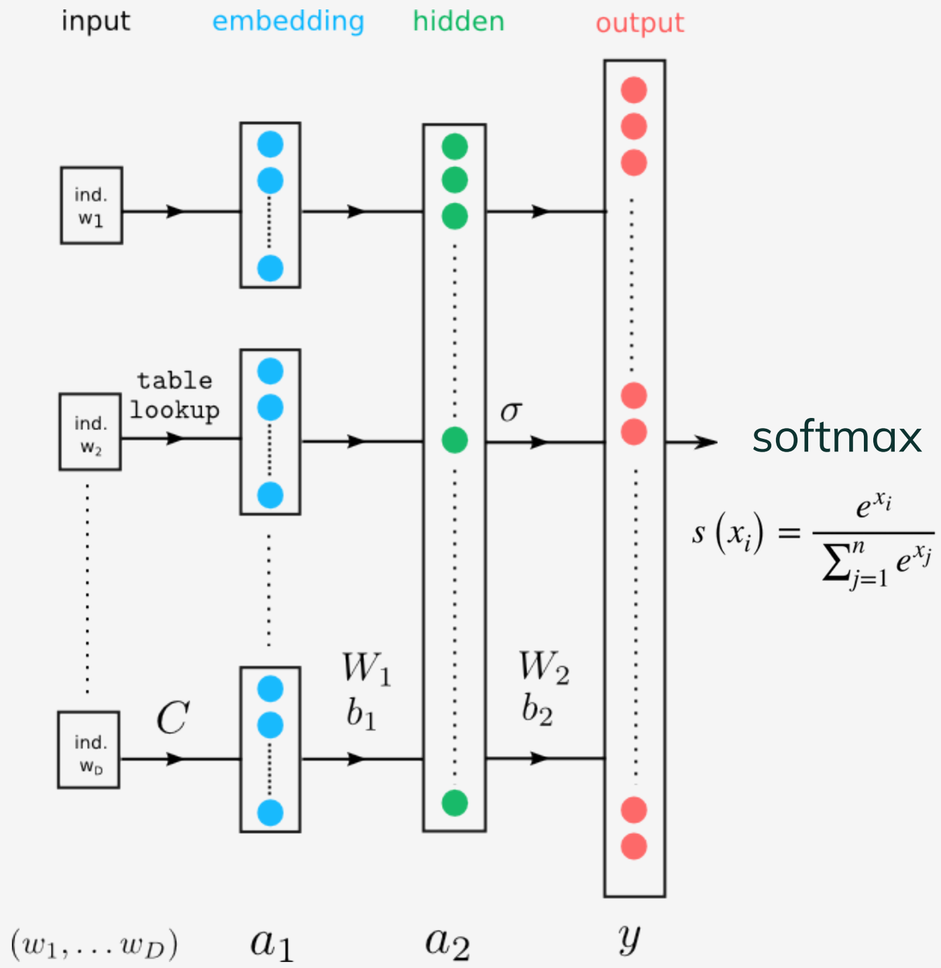
- Use pre-trained models from Hugging Face Transformers.
- Implement a summarization API that can be called from the frontend.

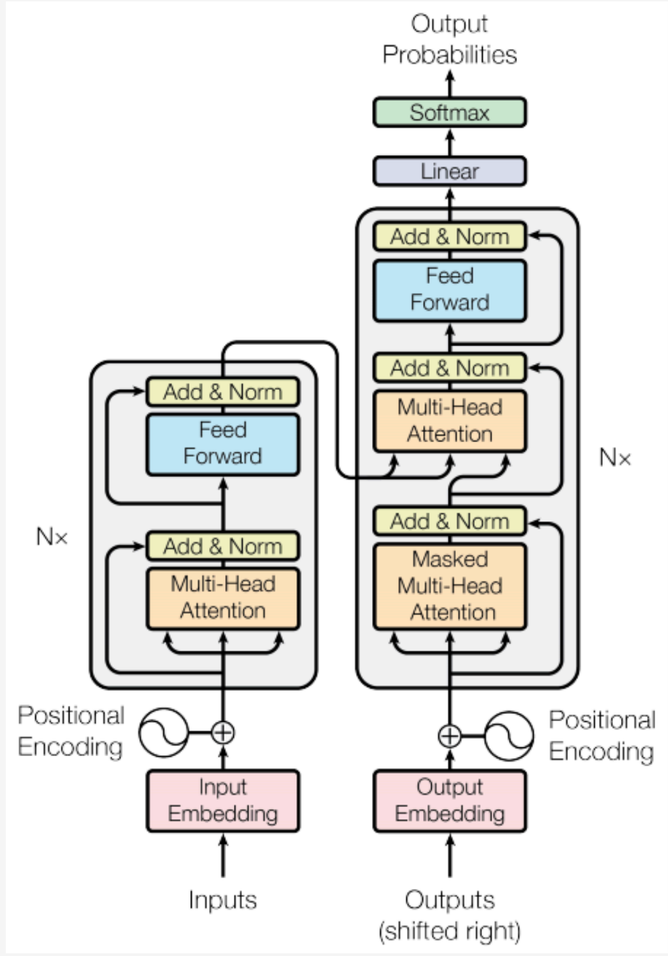


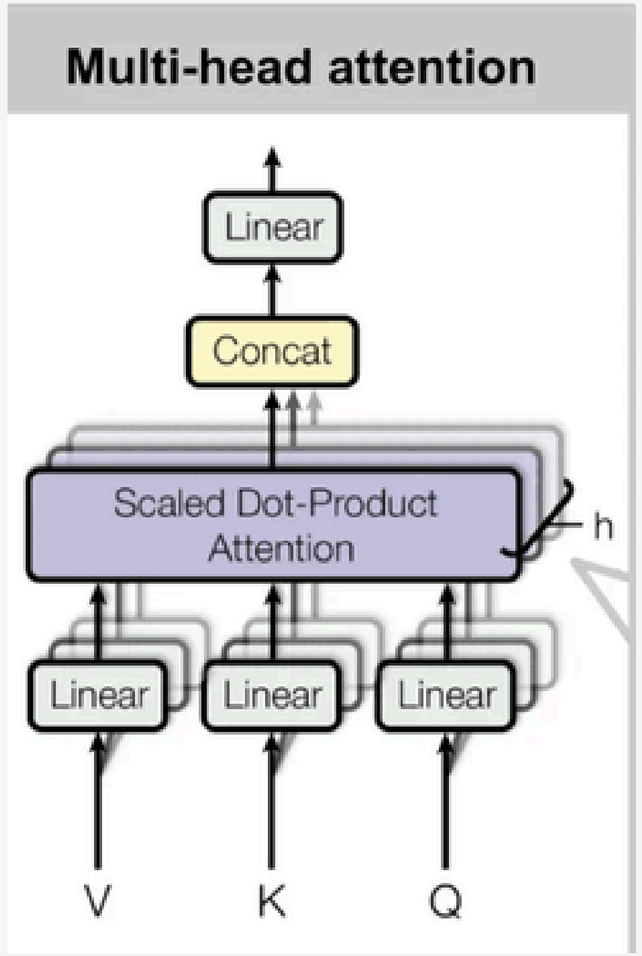
4. Multilingual Support

- Tokenization and splitting into sentences to prepare for translation
- Uses transformer-based models (e.g., MarianMT or OpenNMT) pre-trained on multilingual datasets to translate text
- Applies the attention mechanism from the Transformer architecture to align and generate translated words contextually
- Fine-tuning on specific language pairs or domain-specific data to improve translation accuracy

4. Multilingual Support





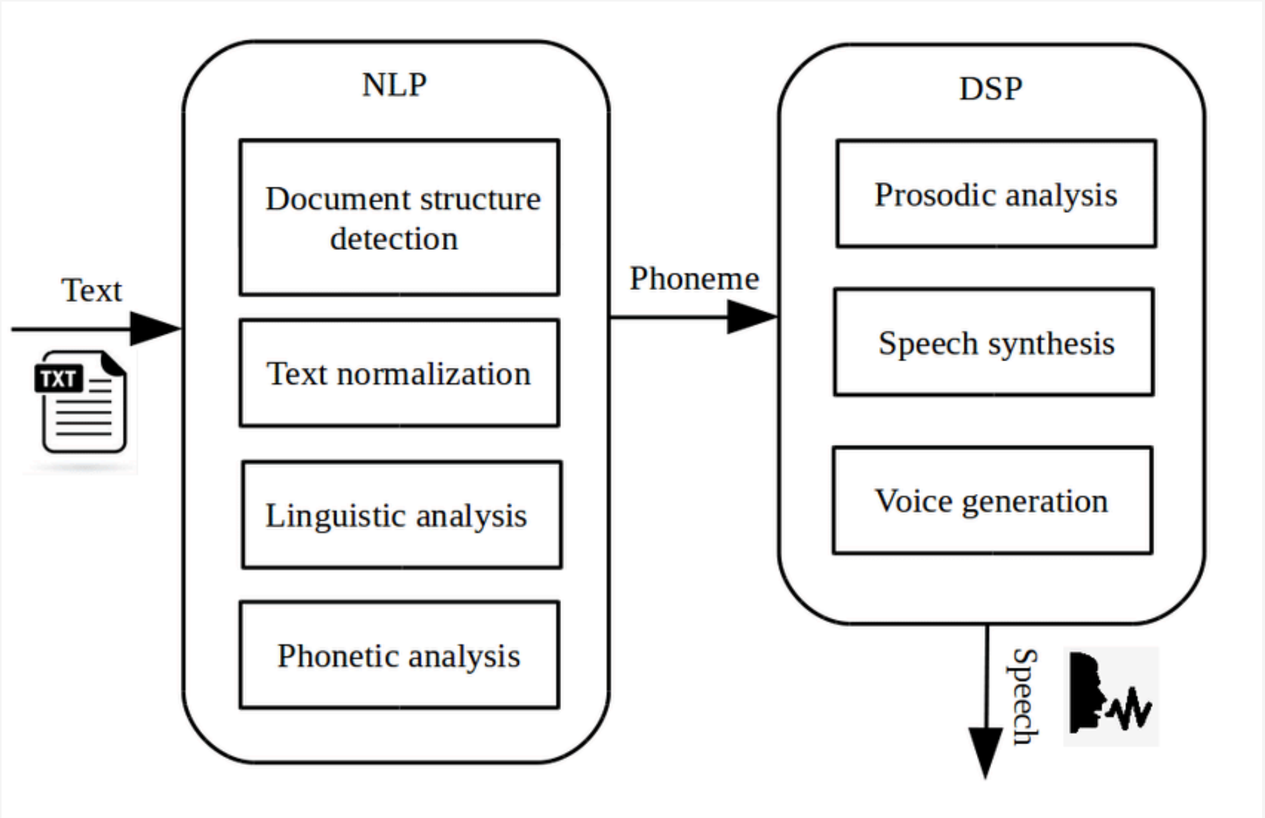


5. Text-to-Speech

- Text Normalization: Converts text into a phonetic representation, handling abbreviations, numbers, and symbols.
- Linguistic Analysis: Uses NLP models to determine the prosody (intonation, rhythm, and stress) for natural-sounding speech.
- Utilize models like Tacotron 2 or FastSpeech, which convert the phonetic sequence into spectrograms that represent the audio waveform.
- Apply a neural vocoder like WaveNet or HiFi-GAN to generate high-fidelity audio from the spectrogram.

5. Text-to-Speech

- Provide options for different languages and voices
- Use Google Text-to-Speech API for reading out notes or chatbot responses



5. Speech-to-Text

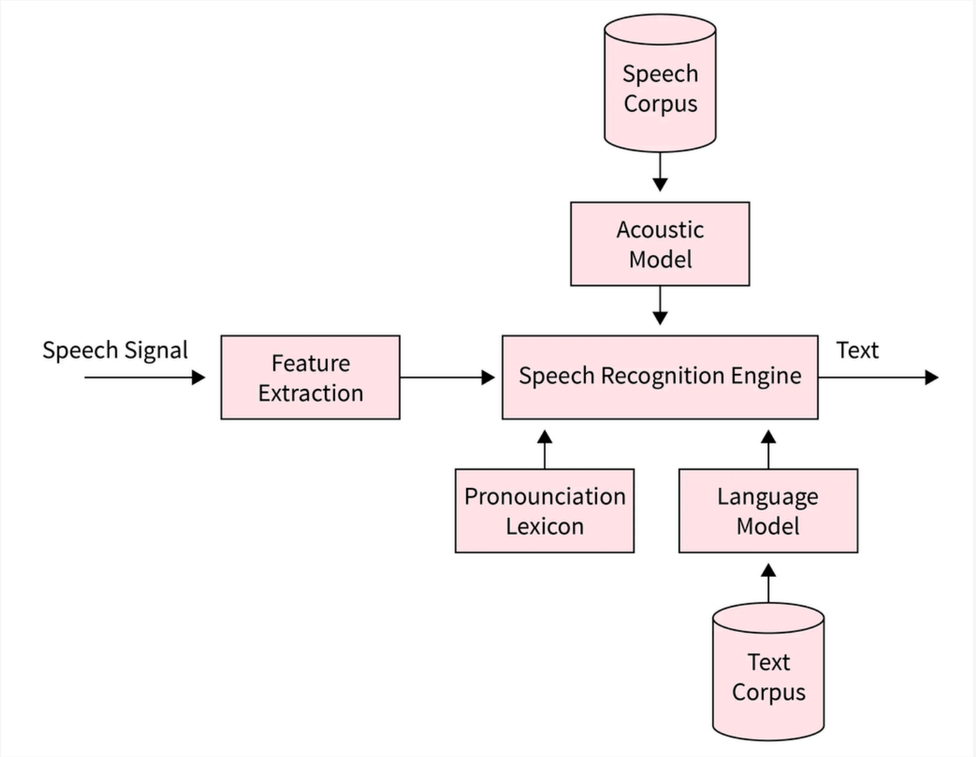
- Noise Reduction: Applies filters (e.g., spectral gating) to remove background noise and enhance the clarity of the speech signal
- Converts audio into Mel-spectrograms or Mel-Frequency Cepstral Coefficients (MFCCs), which capture the spectral characteristics of the audio for further processing
- Uses Recurrent Neural Networks (RNNs) or Transformer-based architectures like Wav2Vec 2.0 for sequence-to-sequence conversion, translating the audio features into phonetic sequences

5. Speech-to-Text

- Fine-tuned models recognize various languages and accents, improving accuracy in multilingual contexts
- Combine the acoustic model output with a language model (e.g., an LSTM-based model) to predict and format words correctly, enhancing transcription quality
- Apply punctuation and capitalization using NLP models, providing a clean, formatted text output.

5. Speech-to-Text

- Use Google Speech-to-Text API for converting spoken input to text
- Implement a feature allowing users to dictate notes or commands



Assumptions

- Users are expected to provide clear, high-resolution images for the OCR module to effectively extract text
- Access to reliable language models and translation algorithms that can handle a wide variety of languages and dialects with minimal errors
- A consistent and stable internet connection is assumed for smooth interaction with the backend and AI modules
- The AI-powered chatbot and NLP systems are expected to improve with continued use, requiring regular updates and model training

Work Breakdown

System Architecture Design
Module Architecture Design

Frontend (Godwin Gino)
Backend (Gautham C Sudheer)

OCR (Godwin Gino)
Text Summarization (Fathima Jennath)
Chatbot (Fathima Jennath)
Translation (Gautham C Sudheer)
Text-to-Speech (Mohammed Basil)
Speech-to-Text (Mohammed Basil)

Database
Testing and Quality Assurance
Deployment
Documentation
Future Enhancements
Scalability

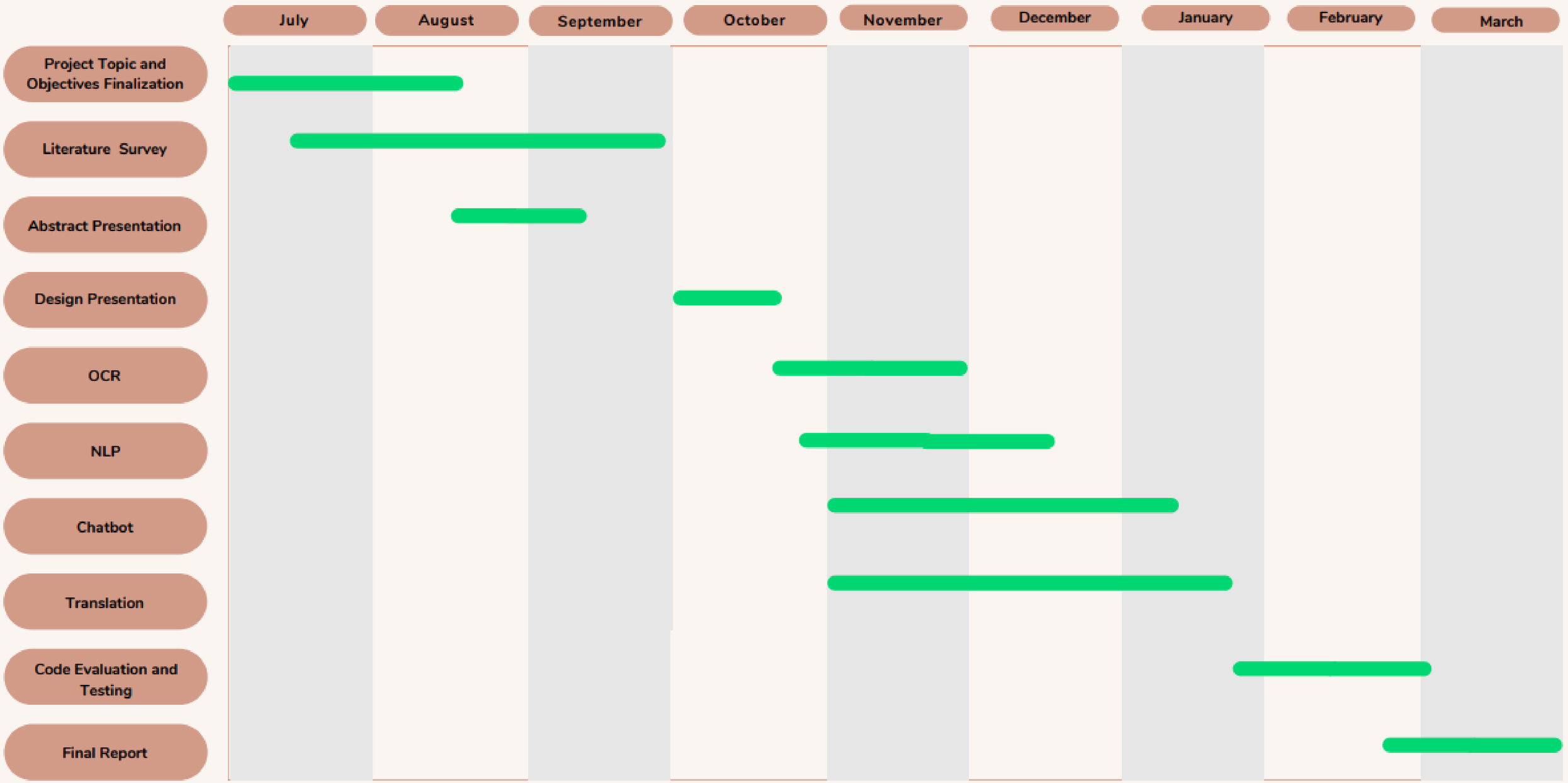
Requirements

- Frontend Development: React.js and React Native
- Backend Development: Python and Django
- OCR and Image Processing: Tesseract OCR, OpenCV, and Pillow (PIL)
- NLP and Machine Learning: spaCy, NLTK, and Transformers
- Chatbot and Conversational AI: Dialogflow and Rasa
- Translation Services: Google Cloud and/or Microsoft Translator API
- Speech Processing: Google Cloud Speech-to-Text API and Text-to-Speech API

Requirements

- Processor: Intel Core i5 or AMD Ryzen 5 (minimum); Intel Core i7 or AMD Ryzen 7 (recommended)
- RAM: 8 GB (minimum); 16 GB or more (recommended)
- Storage: 256 GB SSD (minimum); 512 GB SSD or higher (recommended)
- Graphics Card: Integrated graphics are sufficient for development; dedicated GPU recommended for intensive tasks like machine learning.
- Operating System: Windows 10 or 11, macOS, or a Linux distribution.

Gantt Chart


July
August
September
October
November
December
January
February
March
Project Topic and Objectives Finalization
Literature Survey
Abstract Presentation
Design Presentation
OCR
NLP
Chatbot
Translation
Code Evaluation and Testing
Final Report

BUDGET

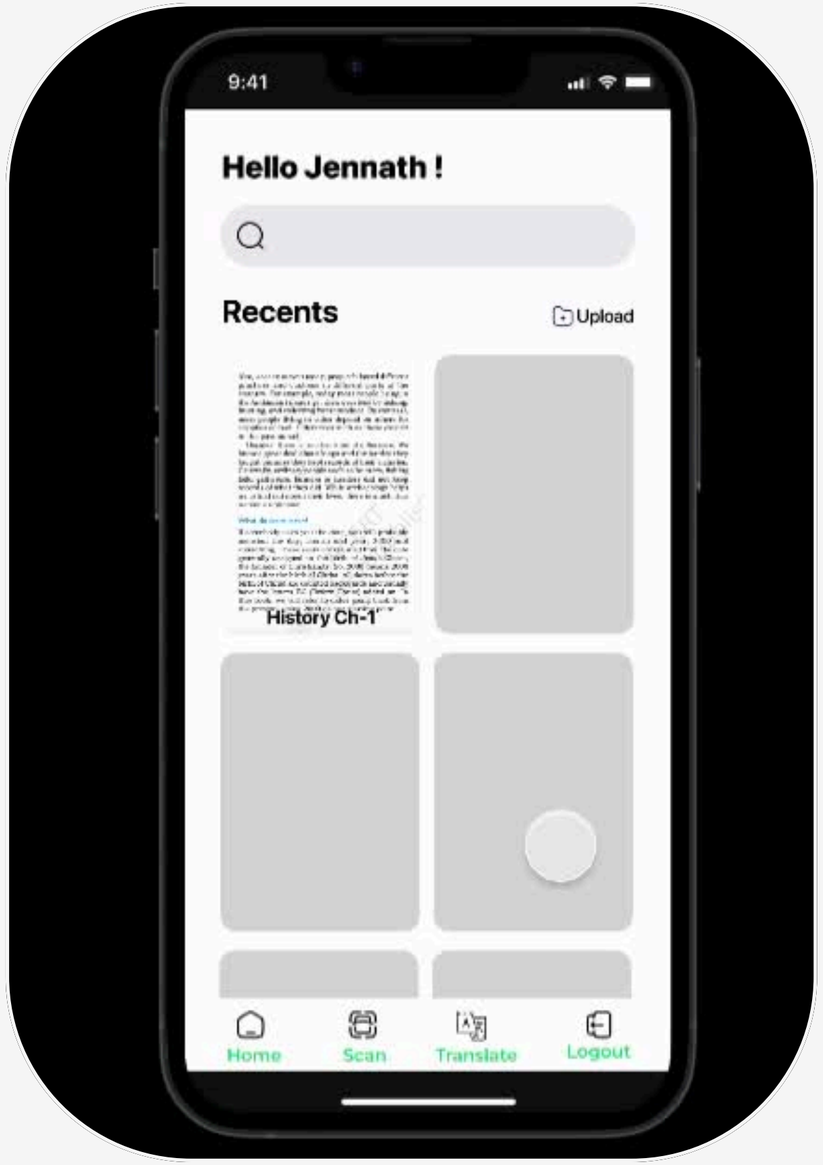
Sl.No.	Items	Amount (Rs.)
1	HIRING OF EXPERTISE / HUMAN RESOURCES	4000
2	SOFTWARE/SERVICES	4000
3	DATA COLLECTION	2000
4	PUBLICATION CHARGES	2000
5	TRAVEL/CONTINGENCY	1000
6	OTHER	1000

Risk and Challenges

- Accurately extracting text from diverse input types and understanding technical language, leading to unreliable outputs.
- Ensuring high translation accuracy across various languages can be challenging, especially with diverse linguistic structures.
- Efficiently handle high volumes of data and concurrent requests without performance degradation, requiring a scalable architecture.
- AI-powered features might demand significant computational resources, leading to potential high operational costs.

Expected Output

- Users can upload text images and receive accurate, editable digital text for easier study and organization.
- An AI-powered chatbot providing real-time assistance by answering questions, generating summaries, and helping users navigate their notes conversationally.
- Robust translation capabilities, enabling users to access educational materials in preferred languages, enhancing inclusivity.
- Utilize speech recognition for note-taking and text-to-speech functionality to listen to notes, promoting accessibility and varied learning methods.



01 Text Recognition

02 Text Extraction

03 Note Organization

04 Multilingual Support

05 Interactive Chatbot

06 Summerization

07 Speech-to-Text

08 Text-to-Speech

Conclusion

"Synapse" is an AI-powered platform that transforms text images into editable digital content, enhancing the educational experience through advanced OCR, NLP techniques, and interactive features.

- Converts printed and handwritten content into editable digital formats, enhancing accessibility and organization
- Provides real-time, interactive management of notes through conversational AI, improving user engagement and efficiency.
- Text summarization, speech processing, and efficient note organization to support diverse learning needs and streamline educational practices.

References

- Dabre, R., Nakazawa, T., & Kunchukuttan, A. (2018). A Voyage on Neural Machine Translation for Indic Languages. Presented at the 6th International Conference on Asian Language Processing (IALP). IEEE.
- Jayan, P., & Poornima, G. (2020). Chatbot Using Deep Learning. Proceedings of the International Conference on Innovative Computing and Communications. Springer, Singapore.
- Ayyalasomayajula, S., & Kolluri, S. S. (2017). CNN-based Optical Character Recognition and Applications. IEEE International Conference on Intelligent Sustainable Systems (ICISS).

References

- Kumar, P. (2019). A Review on CNN and its Applications in Image Processing. Journal of Advances in Mathematics and Computer Science, 29(5), 1-9.
- Wang, J. (2023). A Study of The OCR Development History and Directions of Development. Highlights in Science, Engineering, and Technology, 72, 409-415. DOI: 10.54097/bm665j77. Licensed under CC BY-NC 4.0.

Thank You

DO YOU HAVE ANY QUESTIONS?

SYNAPSE

Gautham C Sudheer | Fathima Jennath N K
Godwin Gino | Mohammed Basil