

FINAL **Synapse**

WHERE AI MEETS EDUCATION

PRESENTATION

Dr. Mary Priya Sebastian

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

CONTENT

01	PROBLEM DEFINITION
02	PURPOSE AND NEED
03	OBJECTIVES
04	LITERATURE SURVEY
05	ARCHITECTURE DIAGRAM
06	MODULE DESCRIPTION

07	GCP VS AWS
08	REQUIREMENTS
09	RISK AND CHALLENGES
10	OUTPUT
11	CONCLUSION
12	REFERENCES

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 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 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

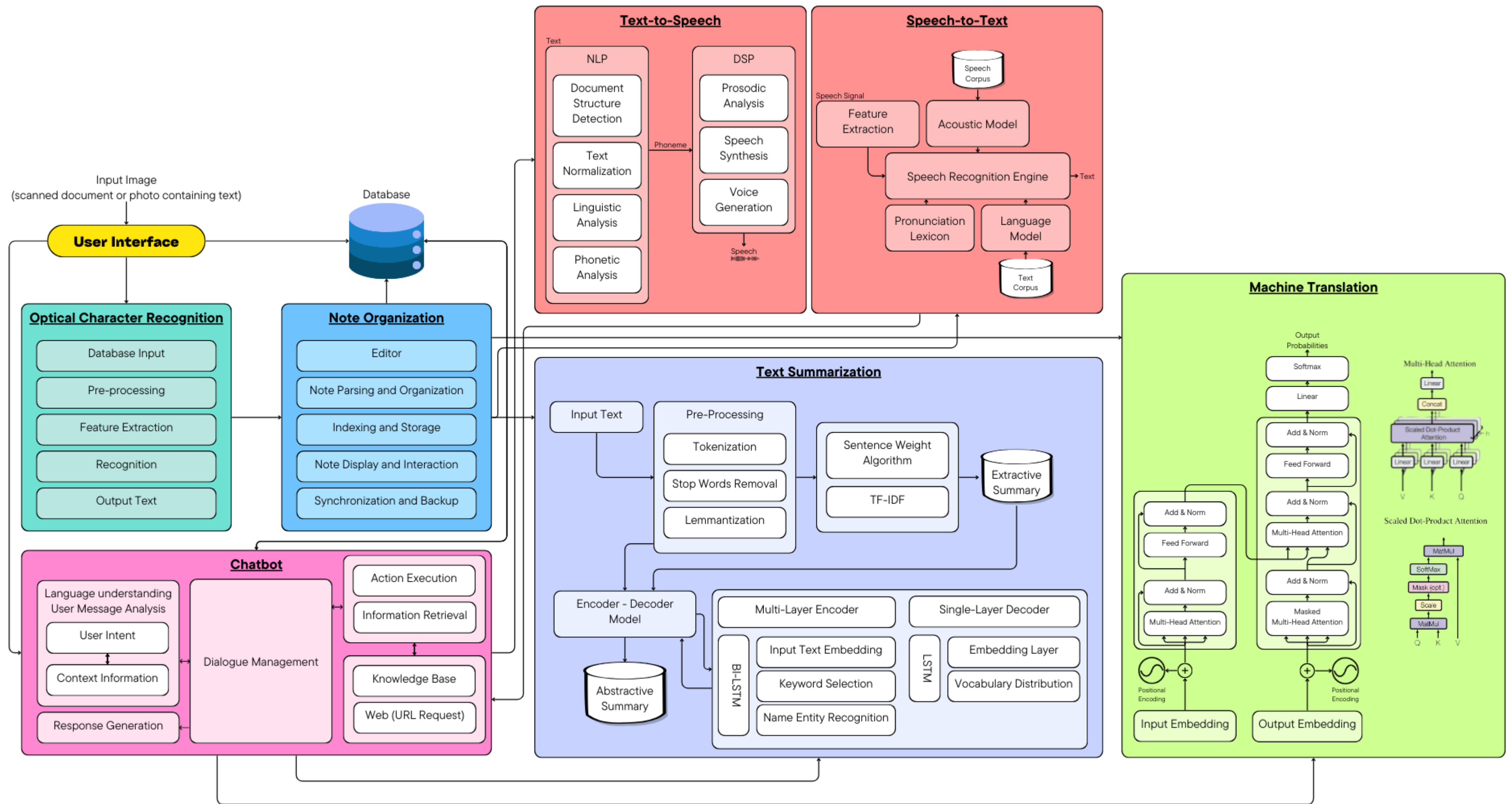
Multilingual and Text-to-Speech, ensuring the platform is adaptable to users from diverse linguistic backgrounds and varying learning needs.

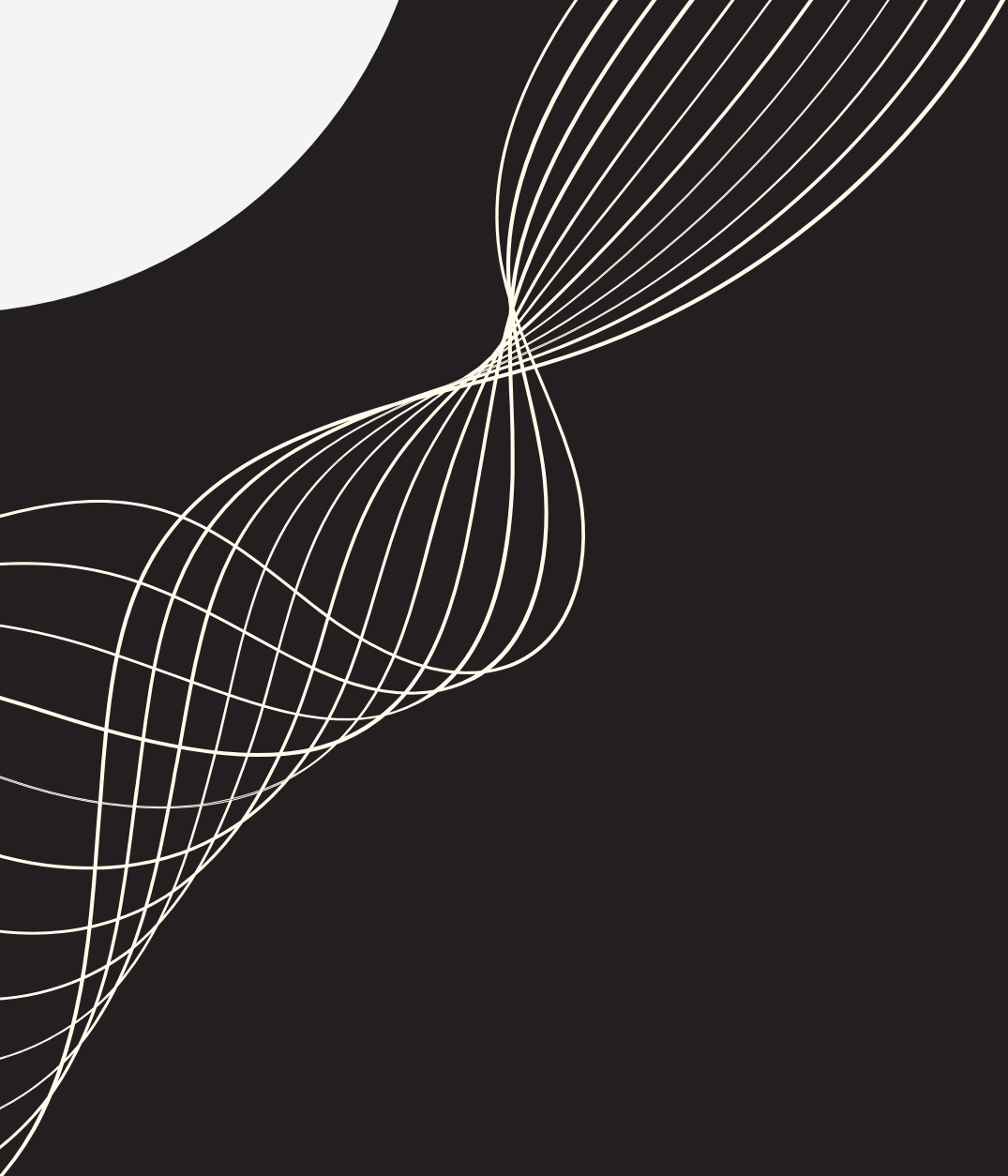
Literature Survey

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.

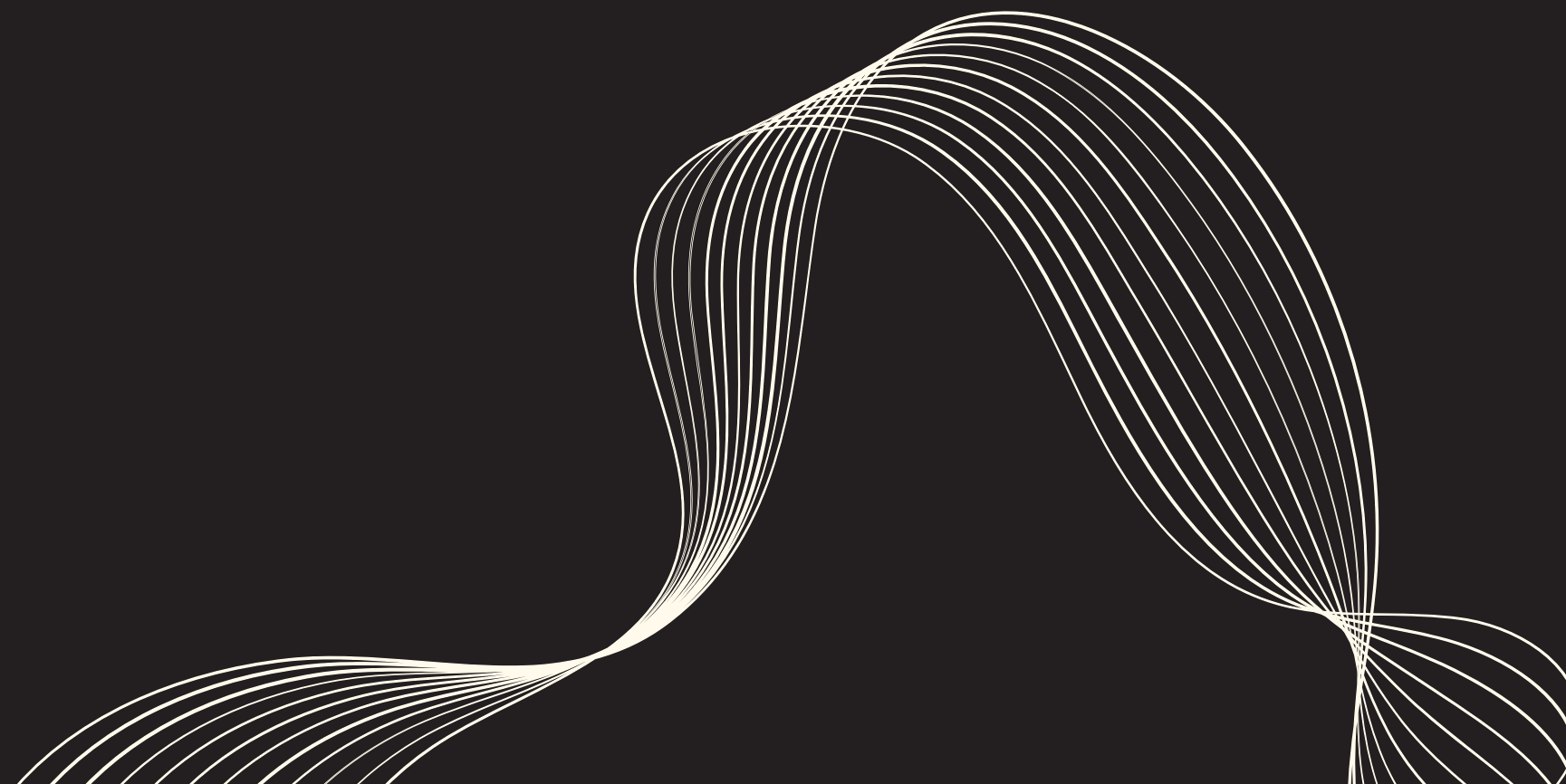
Literature Survey

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





MODULE DESCRIPTION

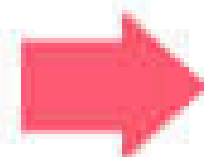




**DOCUMENT
SCAN**



**SCANNED
IMAGE FILE**



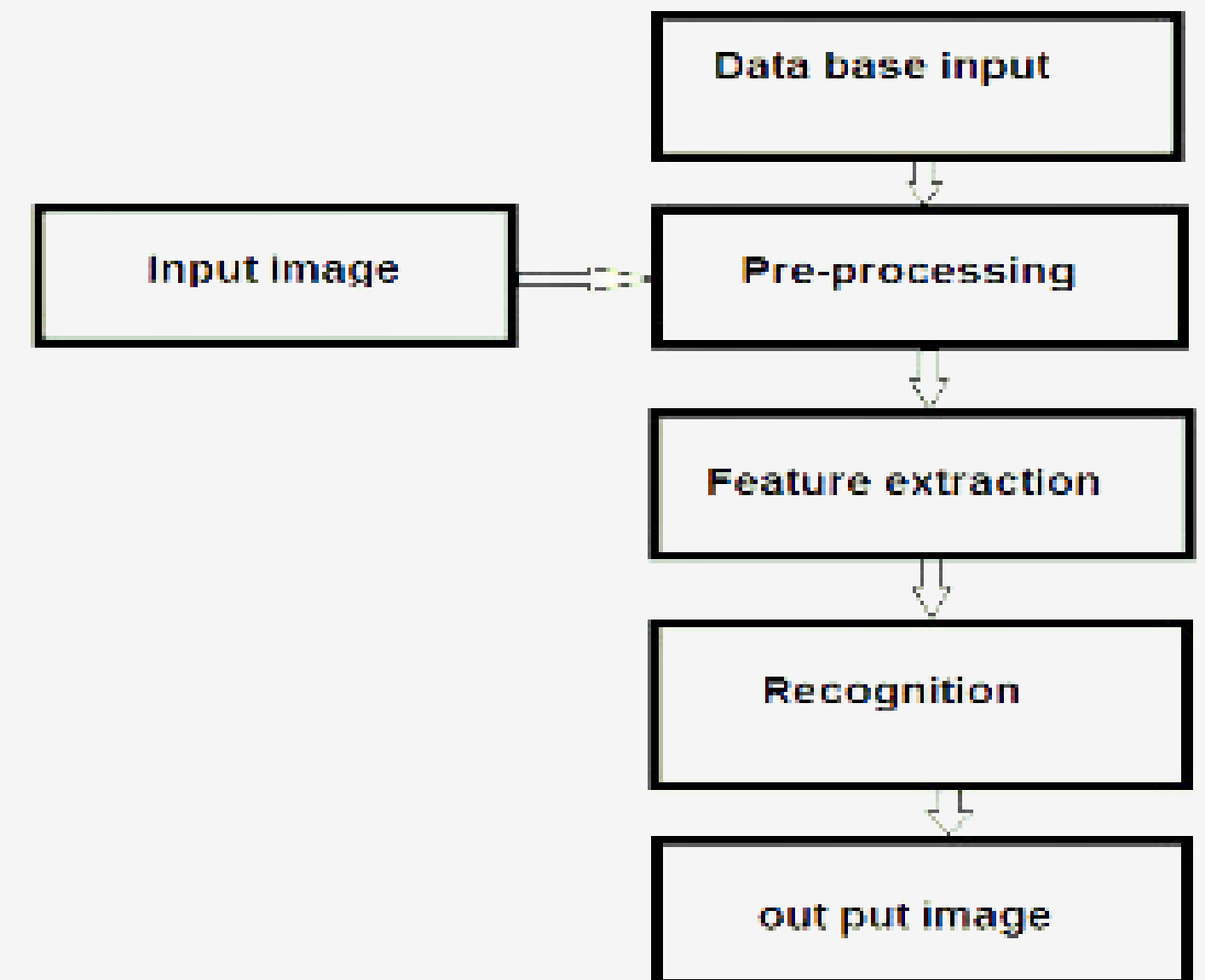
OCR
(Optical Character
Recognition)



**TEXT
DOCUMENT**

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



1. OCR & Text Extraction

- Preprocessing (Enhancing Image Quality)
 - Grayscale Conversion: Reduces complexity, improves contrast.
 - Binarization (Otsu's Thresholding): Converts grayscale to binary for better text separation.
 - Noise Reduction: Gaussian blur/median filters remove distortions.
 - Skew Correction (Hough Transform): Detects & corrects text misalignment.

1. OCR & Text Extraction

- Text Detection
 - EAST (Efficient & Accurate Scene Text Detector): Uses deep learning to locate text regions in images.
- Text Recognition
 - CRNN (Convolutional Recurrent Neural Network):
 - CNN extracts features.
 - LSTM models sequential text for accurate OCR.

OCR System	Accuracy (%)	Character Error Rate (%)	Remarks
AWS Textract	99.3	1.3	Top performer when excluding outlier cases.
Tesseract	98.9	3.6	Open-source; performs well with clear, high-contrast text.
Google Cloud Vision OCR	98.0	5.3	High accuracy across various document types.
Microsoft Azure OCR	87.0	N/A	Struggles with handwritten and complex layouts.

Evaluated using IIIT 5K-word Dataset, Contains 5,000 images of cropped word instances from scene texts and born-digital images, each associated with a 50-word and a 1,000-word lexicon.

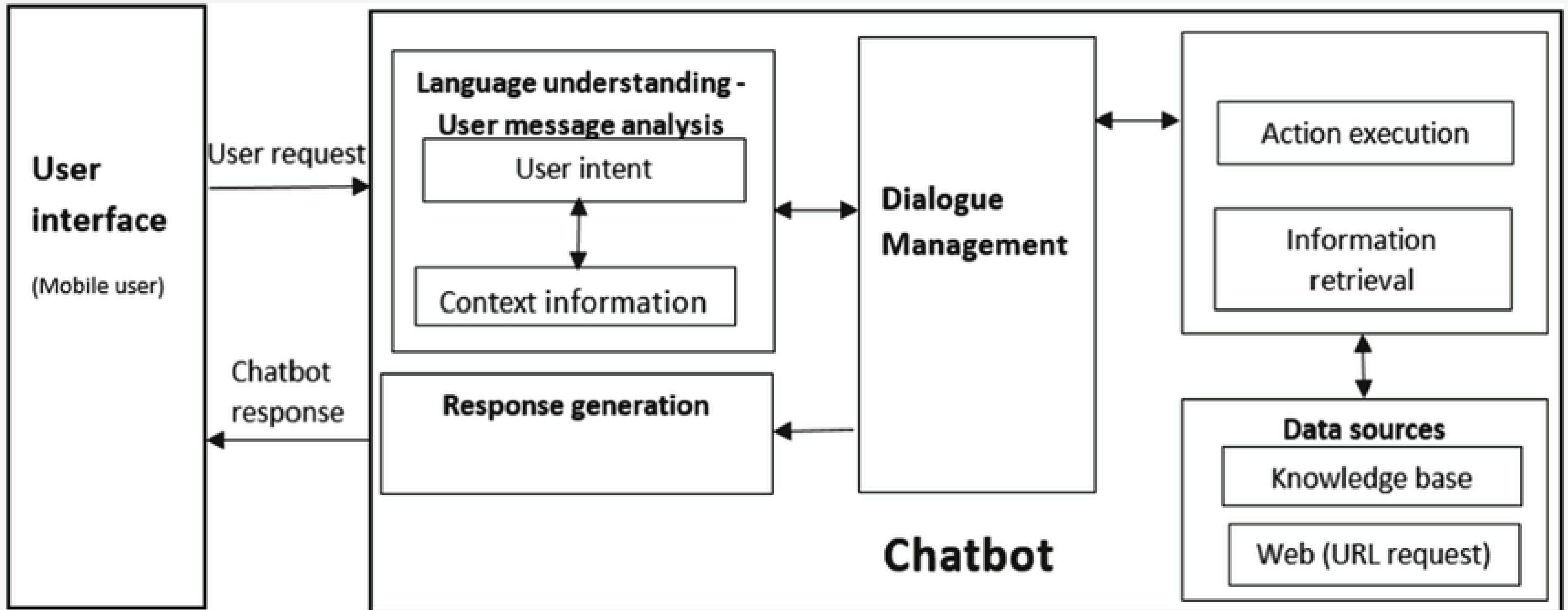
OCR Framework	Word Error Rate (%)	Character Error Rate (%)	Precision (%)
Pytesseract	16.7	3.6	92.3
PyOCR	23.5	7.8	90.6
EasyOCR	34.3	19.4	83.4



2. AI-Powered Chatbot

- Natural Language Understanding (NLU)
 - Tokenization & Named Entity Recognition (NER): Identifies key entities using NLP tools (SpaCy, Hugging Face).
 - BERT/DistilBERT: Captures context for better query understanding.
- Intent Recognition
 - Classification Model (BERT/RNN-based): Determines user intent.
 - Entity & Context Extraction: Ensures precise responses.
- Response Generation
 - GPT/Dialogflow: Generates context-aware replies.

2. AI-Powered Chatbot



Chat With Your Document

👋 Hello! I'm SynapseAI, and I'm here to help you understand and analyze the extracted text. Here's what I can do for you:

- Ask me questions about the text
- Request a summary of the content
- Get explanations of specific parts
- Ask for key points or main ideas
- Request analysis or insights
- Click the speaker icon 🗣️ to hear my responses

Feel free to start by asking any question about the text!



Type your message...

Send

Chat With Your Document

What is this about?

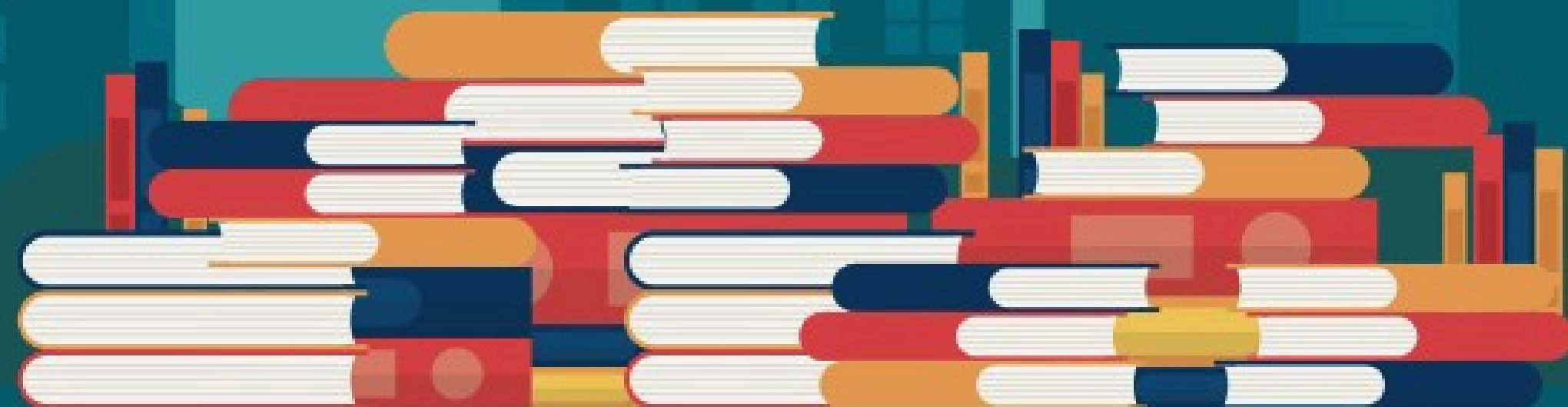
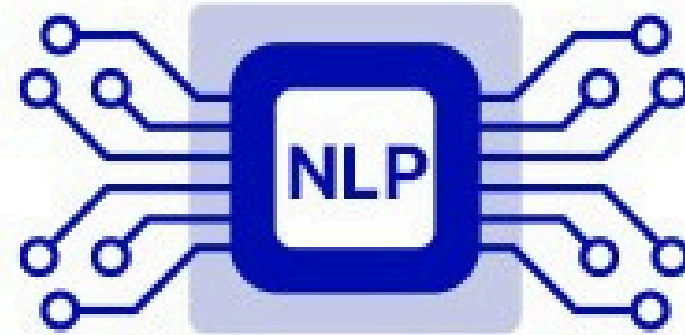
This document is an introduction to plate tectonics. It explains the theory of plate tectonics, describes the three main types of plate boundaries (divergent, convergent, and transform), and details the various geological phenomena resulting from tectonic activity, such as earthquakes, volcanism, mountain building, and ocean basin formation. It also connects plate tectonics to continental drift and discusses the theory's impact on human civilization, including natural disasters, resource distribution, and climate. In short, it's an overview of how the movement of Earth's plates shapes our planet and influences human life.



Type your message...

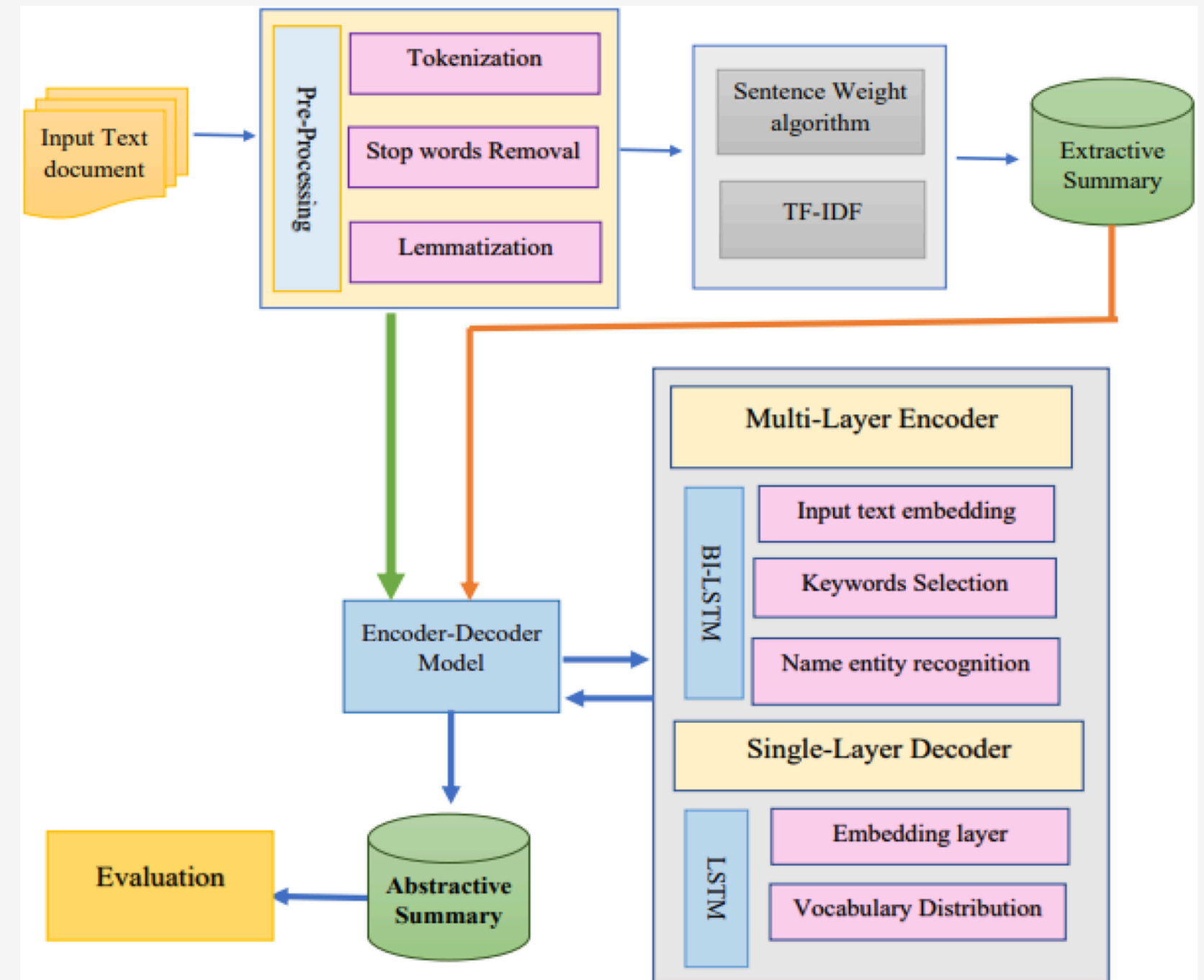
Send

SUMMARY



3. Text Summarization

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



3. Text Summarization

- Preprocessing
 - Tokenization: Splits text into words & sentences.
 - Lemmatization & Stopword Removal: Reduces words to base forms & removes irrelevant words.
- Feature Representation
 - Word Embeddings (Word2Vec, GloVe): Captures word meanings.
 - Contextual Embeddings (BERT): Understands word meaning in context.

Chatbot System	General Knowledge (MMLU)	Context Window	Multimodal Capabilities
Gemini	90.0	Up to 1 million tokens	Text, images, video, voice
DeepSeek	90.8	Up to 128K tokens	Text-based only
ChatGPT (GPT-4o)	88.3	Up to 128K tokens	Text and images

Evaluated using the Massive Multitask Language Understanding (MMLU) benchmark. MMLU comprises approximately 16,000 multiple-choice questions spanning 57 subjects, including mathematics, history, computer science, law, and more.

Chatbot System	Natural Language Understanding (NLU)	Integration Capabilities	Pricing	Supported Languages
Amazon Lex	Yes	AWS services; platforms like Facebook Messenger, Slack, Twilio, Kik	\$0.004 per speech request; \$0.00075 per text request	Primarily English
Google Chat API	No	Google Workspace services; enhances Google Chat functionalities	Varies based on usage; aligned with Google Cloud's pricing models	Multiple languages, depending on implementation
Meta Messenger API	No	Facebook Messenger; connects to other services via webhooks and APIs	Varies based on usage;	Multiple languages, depending on implementation

Chat With Your Document

Can you summarize this document?

This document explains plate tectonics, the theory that Earth's lithosphere is divided into moving plates. These plates interact at three main boundaries: divergent (plates move apart, creating new crust), convergent (plates collide, causing subduction or mountain building), and transform (plates slide past each other, causing earthquakes). Plate tectonics explains various geological phenomena, including earthquakes, volcanism, mountain building, and ocean basin formation. It also supports the continental drift theory, proposing that continents were once joined in a supercontinent called Pangaea. Understanding tectonics is crucial for predicting and mitigating natural disasters, managing resources, and comprehending Earth's history and evolution.



Type your message...

Send

Chat With Your Document

ഇത് എന്തിനെക്കുറിച്ചാണ്?

ഈ രേഖ മലയാളം വ്യാകരണത്തെക്കുറിച്ചാണ്. അതിൽ നാമങ്ങൾ, ക്രിയാവിശേഷണങ്ങൾ, അവ്യയങ്ങൾ, സമാസങ്ങൾ, സന്ധികൾ, കാലഭേദങ്ങൾ എന്നിവയെക്കുറിച്ചുള്ള വിവരങ്ങളുണ്ട്.



Type your message...

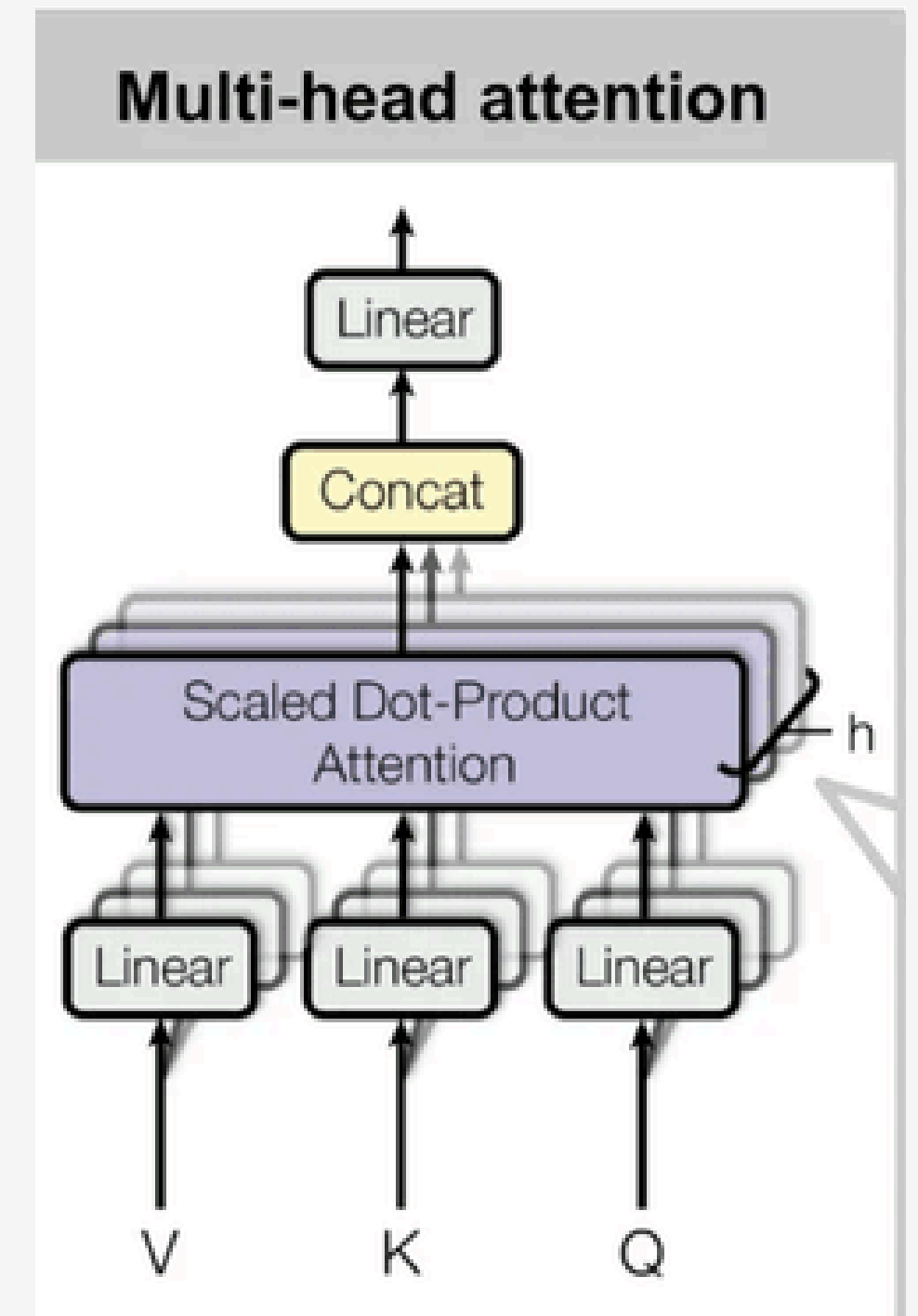
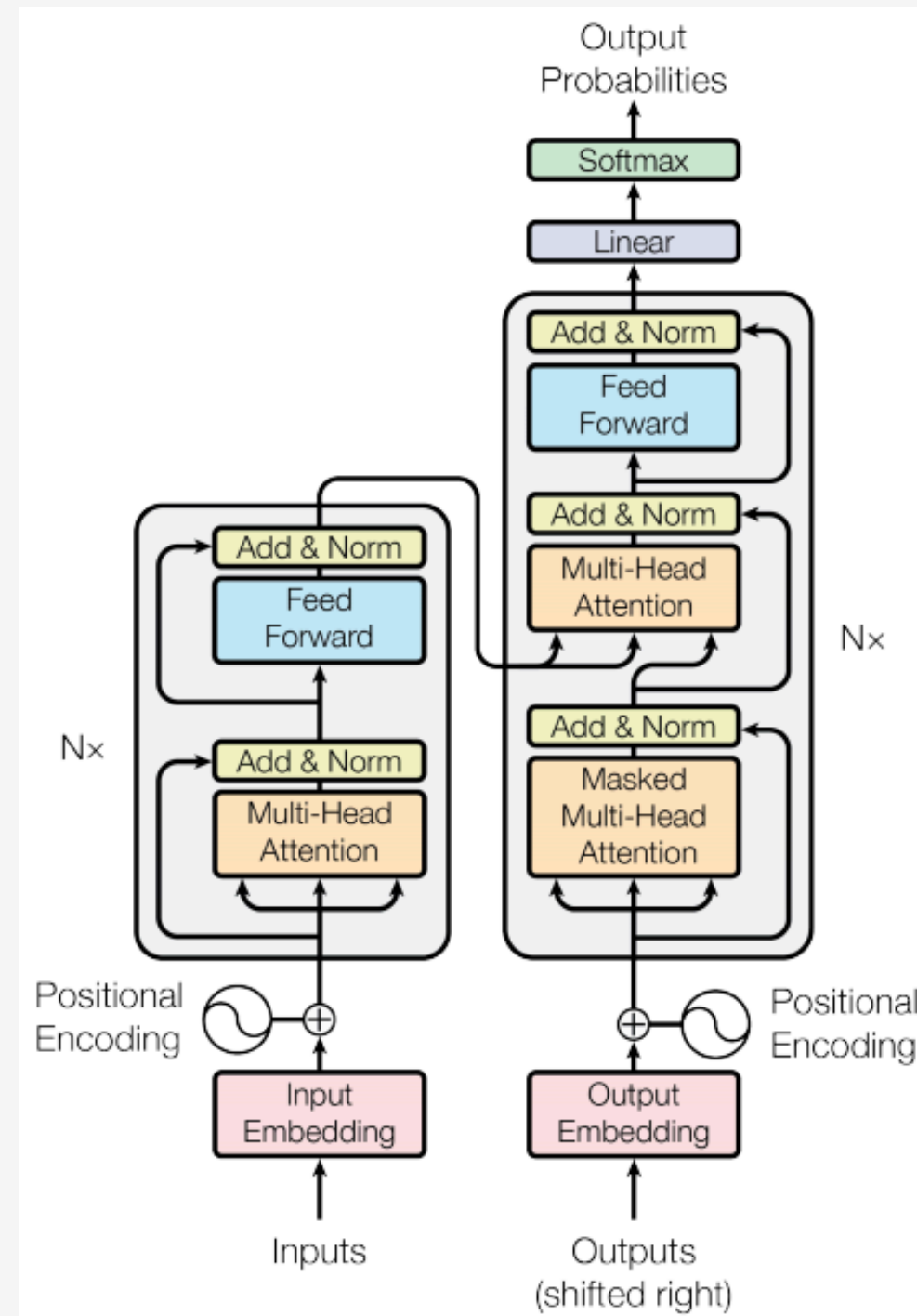
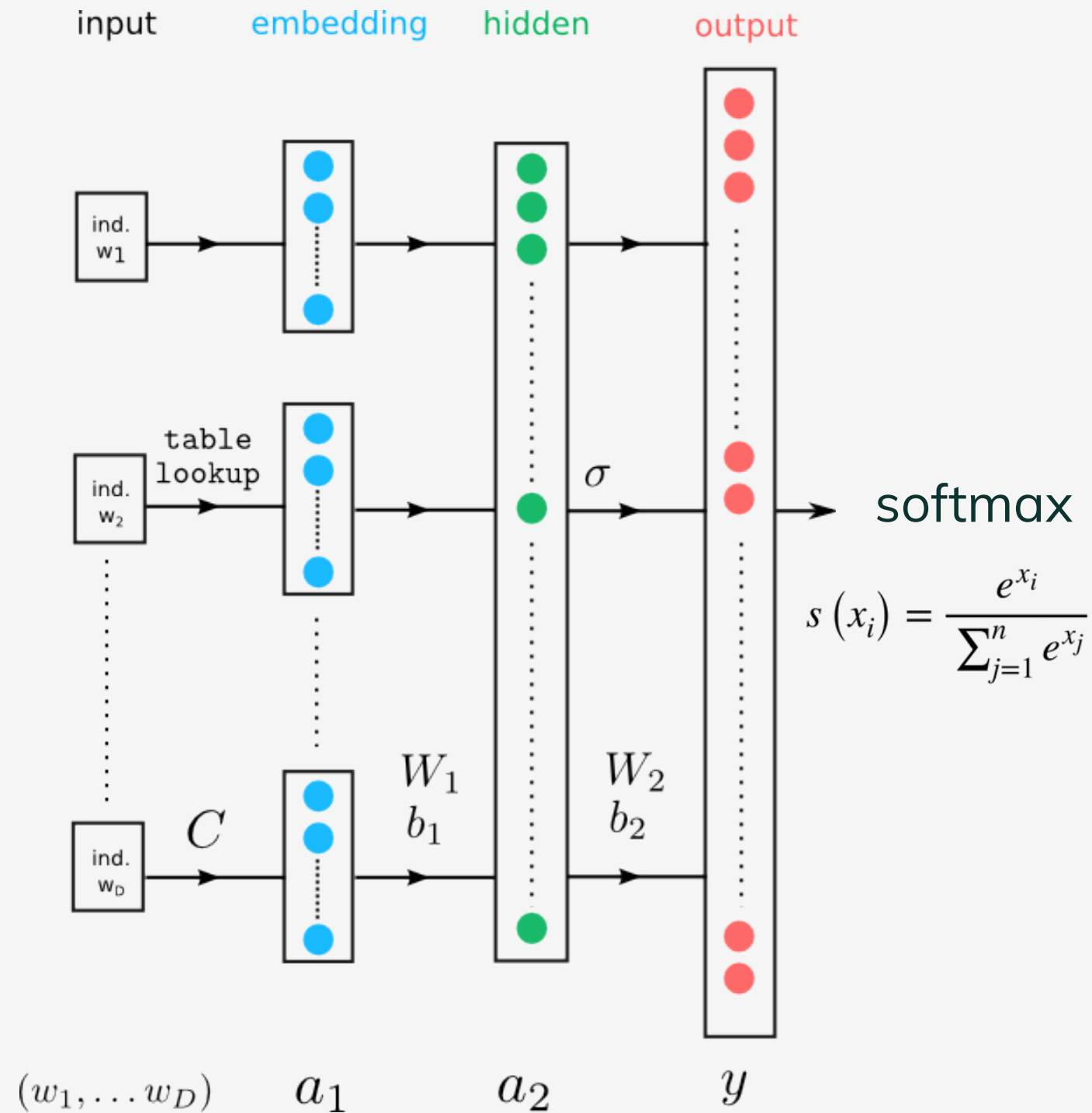
Send



4. Multilingual Support

- Preprocessing
 - Tokenization & Sentence Splitting: Prepares text for translation.
- Translation Model
 - Transformer-based Models: Leverages multilingual datasets.
 - Attention Mechanism: Aligns & generates context-aware translations.
- Optimization
 - Fine-tuning: Enhances accuracy for specific languages & domains.

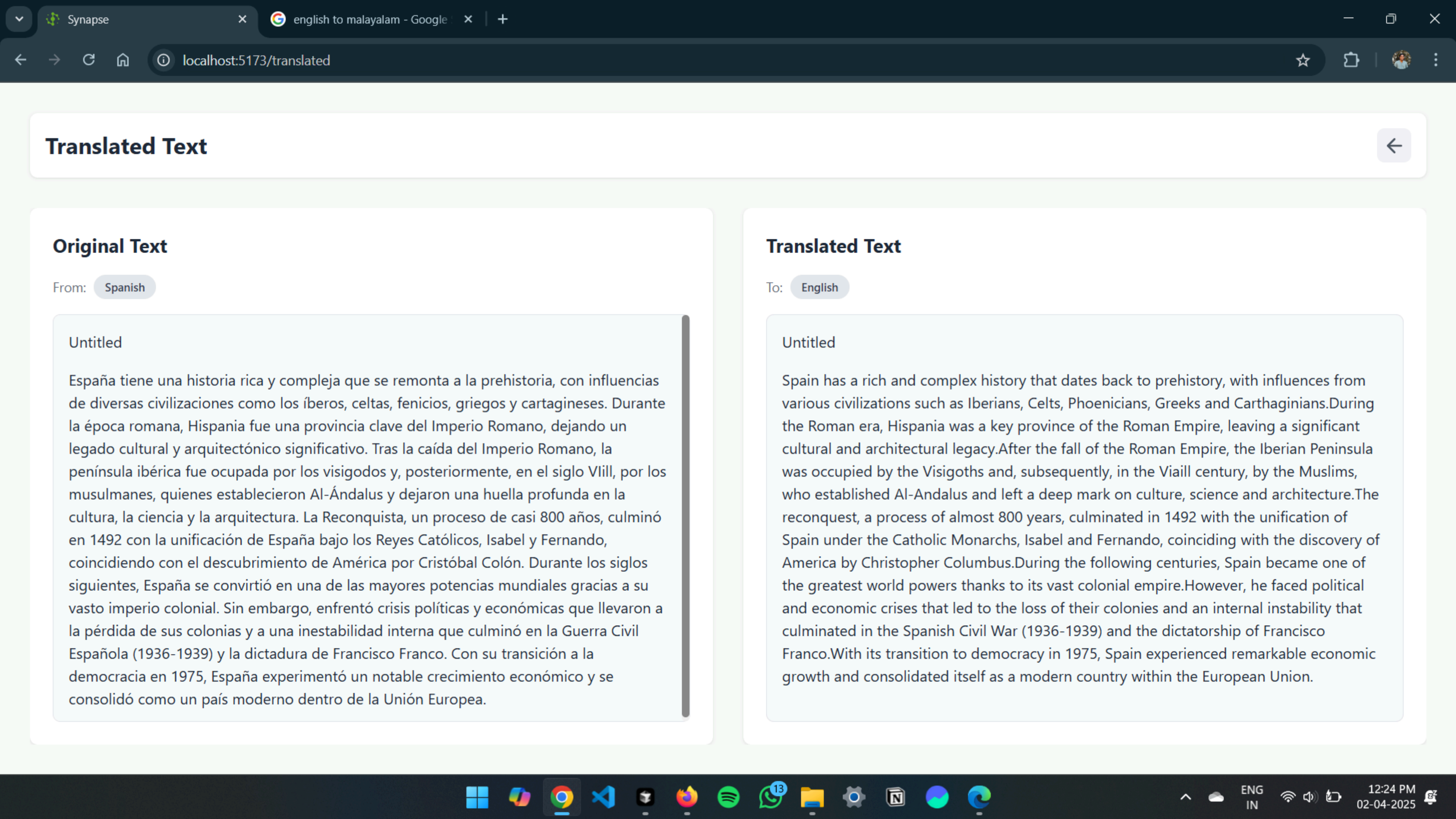
4. Multilingual Support



Approach	Domain	BLEU Score	Accuracy	Notes
NMT	Spoken	0.4273	90.42%	Captures broader context; requires substantial data and computational resources.
SMT	Spoken	0.3671	89.58%	Learns from bilingual corpora; may misinterpret context without sufficient data.
RBMT	Spoken	0.2570	84.17%	Effective with structured language; struggles with idiomatic expressions.

Translation Service	Average Translation Quality Score	Processing Speed (seconds)	Supported Languages
DeepL	8.38	0.51	31
Google Translate	7.90	0.22	133
Microsoft Translator	7.77	0.26	100+
Amazon Translate	N/A	0.33	75

Evaluated using WMT (Workshop on Machine Translation) Datasets, and IWSLT (International Workshop on Spoken Language Translation) Datasets, annual datasets provided by the WMT and IWSLT conference for shared tasks in MT, covering multiple language pairs and domains.



Translated Text



Original Text

From: Spanish

Untitled

España tiene una historia rica y compleja que se remonta a la prehistoria, con influencias de diversas civilizaciones como los íberos, celtas, fenicios, griegos y cartagineses. Durante la época romana, Hispania fue una provincia clave del Imperio Romano, dejando un legado cultural y arquitectónico significativo. Tras la caída del Imperio Romano, la península ibérica fue ocupada por los visigodos y, posteriormente, en el siglo VIll, por los musulmanes, quienes establecieron Al-Ándalus y dejaron una huella profunda en la cultura, la ciencia y la arquitectura. La Reconquista, un proceso de casi 800 años, culminó en 1492 con la unificación de España bajo los Reyes Católicos, Isabel y Fernando, coincidiendo con el descubrimiento de América por Cristóbal Colón. Durante los siglos siguientes, España se convirtió en una de las mayores potencias mundiales gracias a su vasto imperio colonial. Sin embargo, enfrentó crisis políticas y económicas que llevaron a la pérdida de sus colonias y a una inestabilidad interna que culminó en la Guerra Civil Española (1936-1939) y la dictadura de Francisco Franco. Con su transición a la democracia en 1975, España experimentó un notable crecimiento económico y se consolidó como un país moderno dentro de la Unión Europea.

Translated Text

To: English

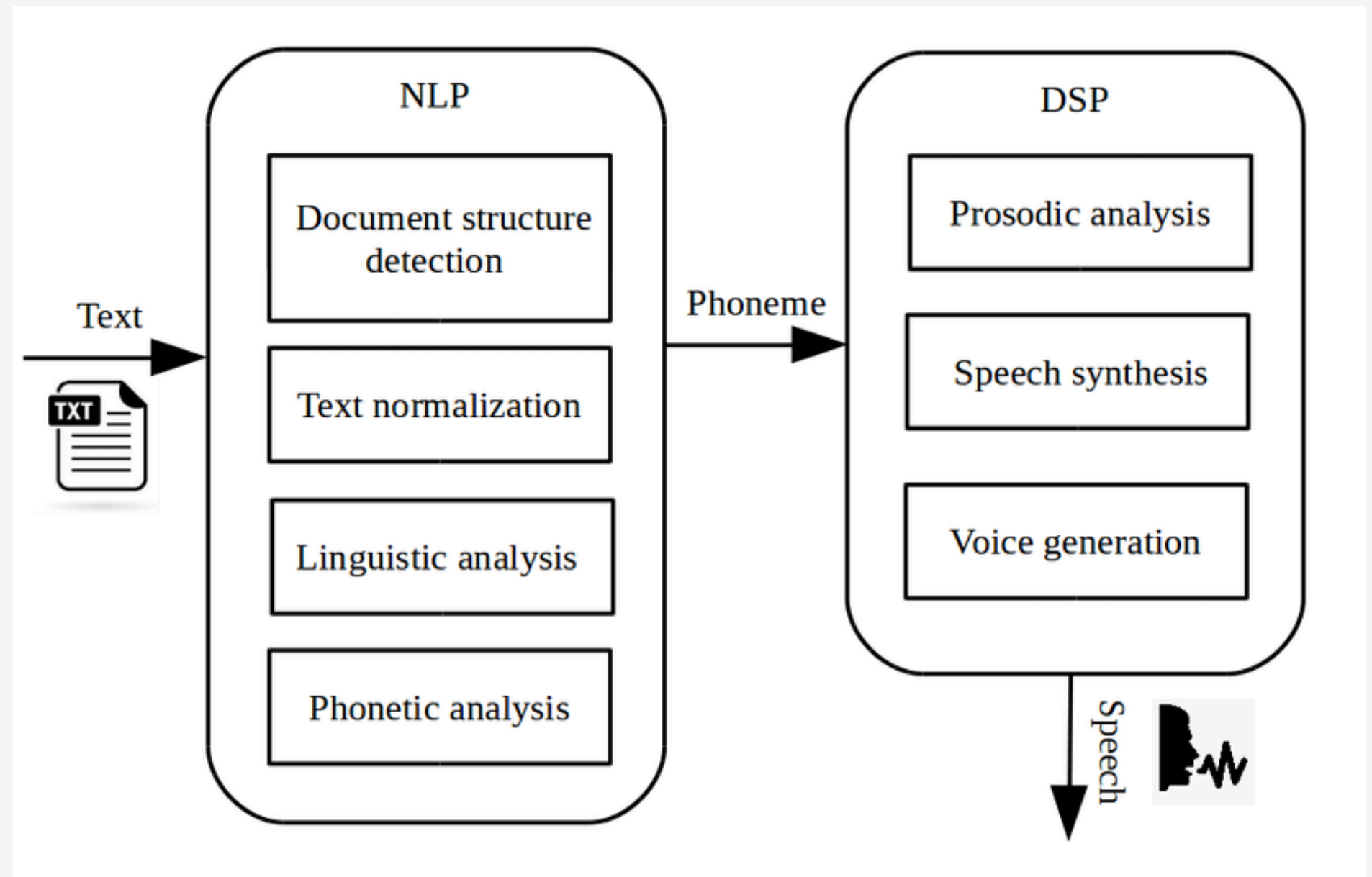
Untitled

Spain has a rich and complex history that dates back to prehistory, with influences from various civilizations such as Iberians, Celts, Phoenicians, Greeks and Carthaginians. During the Roman era, Hispania was a key province of the Roman Empire, leaving a significant cultural and architectural legacy. After the fall of the Roman Empire, the Iberian Peninsula was occupied by the Visigoths and, subsequently, in the VI century, by the Muslims, who established Al-Andalus and left a deep mark on culture, science and architecture. The reconquest, a process of almost 800 years, culminated in 1492 with the unification of Spain under the Catholic Monarchs, Isabel and Fernando, coinciding with the discovery of America by Christopher Columbus. During the following centuries, Spain became one of the greatest world powers thanks to its vast colonial empire. However, he faced political and economic crises that led to the loss of their colonies and an internal instability that culminated in the Spanish Civil War (1936-1939) and the dictatorship of Francisco Franco. With its transition to democracy in 1975, Spain experienced remarkable economic growth and consolidated itself as a modern country within the European Union.



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. Text-to-Speech

- Preprocessing
 - Text Normalization: Converts text into phonetic form (handles abbreviations, numbers, symbols).
- Linguistic Processing
 - Prosody Analysis: NLP models determine intonation, rhythm, and stress for natural speech.

5. Text-to-Speech

- Speech Synthesis
 - Tacotron 2 / FastSpeech: Converts phonetic sequences into spectrograms.
 - Neural Vocoder (WaveNet / HiFi-GAN): Generates high-quality speech from spectrograms.

Service	Number of Languages	Voice Options	Pricing	Key Features
Google Cloud Text-to-Speech	50+	Multiple voices	Usage-based pricing	Customizable voices, integration with Google Cloud services
Microsoft Azure Text to Speech	140+	Multiple voices	Usage-based pricing	Customizable voices, integration with Azure services
Amazon Polly	39+	Multiple voices	Usage-based pricing	based pricing Realistic speech generation, integration with AWS services
Synthesia	130+	Multiple voices	\$22 per month	Video editor, dubbing, subtitles, transcription



Features

Use Cases

Try Now

multilingual support, and natural voice synthesis.

Try It Free →

▶ Watch Demo

Powerful Features

Everything you need to digitize, understand, and interact with your documents



OCR Technology

Transform any text image or handwritten notes into searchable digital content



AI Chat Assistant

Engage with your documents through our intelligent chatbot. Get



Multilingual Support

Break language barriers with support for 100+ languages. Automatically detect and



Text-to-Speech

Listen to your documents with natural-sounding voice synthesis. Multiple voices

Comparing GCP & AWS

NLP Services

Service	Feature	Google Cloud	AWS
OCR	API	Cloud Vision API	Amazon Textract
	Capabilities	Text detection, handwriting recognition, multi-language support	Text extraction, table and form data extraction, handwriting recognition
	Supported Formats	Images (JPEG, PNG, GIF)	Documents (PDF, TIFF), Images (JPEG, PNG)
	Pricing	Based on image count; details at Google Cloud Vision Pricing	Based on pages processed; details at Amazon Textract Pricing

Service	Feature	Google Cloud	AWS
Machine Translation	API	Translation API	Amazon Translate
	Supported Languages	100+	75
	Customization	Glossary support	Custom terminology
	Pricing	\$20 per million characters; details at Google Cloud Translation Pricing	\$15 per million characters; details at Amazon Translate Pricing

Service	Feature	Google Cloud	AWS
Text-to-Speech (TTS)	API	Text-to-Speech	Amazon Polly
	Languages & Voices	50+ languages, multiple voices	41 language variants, 100+ voices
	Features	Voice cloning, pitch and speed control	Per-word timestamps, pitch and speed control
	Pricing	Free tier: 1M characters/month; Standard: \$16 per 1M characters; details at Google Cloud Text-to-Speech Pricing	Free tier: 1M characters/month for 12 months; Standard: \$16 per 1M characters; details at Amazon Polly Pricing

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

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.

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.

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 text-to-speech functionality to listen to notes, promoting accessibility and varied learning methods.

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 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?



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