International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04700
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Veda-VisionGPT-WhereDocuments Speak Your Own Language

Kunal Sapkal,Sanika Kadam,Vinayak Gharge,Rutuja Mulik,Vaibhav Bhosale

Abstract:- **Veda-Vision GPT is a transformative multilingual document processing platform designed to handle the complexities of India's diverse linguistic landscape. Processing and translating printed or scanned materials is difficult in a nation with so many regional scripts and dialects. To improve document accessibility, this platform integrates text extraction (OCR), machine translation, and intelligent search features. More than 15 Indian languages receive assistance, including challenging ones like Sanskrit and Kannada.**

The system properly removes and translates the text from the scanned pages with one click. Users can also ask questions about the document and get a thanksful answer to its AI-Interested Query System. In addition, integrated visual tools help users to know language trends and material insight. It also keeps fairness and inclusivity in focus by reducing language-based barriers.

I. INTRODUCTION

Problem Addressed

Multilingual documents can be difficult to use in the modern world, especially between diverse nations like India. In a country like India, where people speak different types of languages, the documents written in different scripts can be difficult and difficult to understand. Nowadays many systems have issues with complex scripts or scanned images, and many local languages still lack the required translation support. People are difficult to find important details as a result.Our project focuses on solving this by making document reading, translation, and understanding much easier and faster.

Importance of the Problem

For diverse linguistic groups to have fair access to information, efficient document handling is crucial. It makes governance easier, healthcare better, education more accessible, and legal systems stronger. Language barriers continue to impede communication, inclusivity, and development in a variety of sectors in the absence of trustworthy solutions.**Summary of the Implemented System**

Veda-Vision GPT integrates advanced OCR, NMT, and AI-driven question-answering to offer a seamless, multilingual document processing. It supports over 15 Indian languages, processes printed and scanned PDF/Image text, and provides integrated translation capability and intuitive visual analytics for content insights.

Summary of Work

Existing tools like Google Translate and standalone OCR systems offer partial solutions but lack the integration and accuracy required for Indian scripts and low-resource languages. Previous research highlights the limitations of traditional machine translation and search tools, particularly for complex linguistic structures and real-time queries. Veda-Vision GPT overcomes these gaps by combining state-of-the- art AI technologies with domain-specific optimizations.

Summary of Methodology

We started by collecting data from documents written in different Indian languages. Then we trained our own OCR and translation tools and connected them with a smart question-answer system. The platform runs on the cloud and uses visual libraries like Plotly to show insights in a simple way.

Evaluation Results

Evaluation demonstrated high OCR accuracy for Indian scripts, improved translation quality for low- resource languages, Then we trained our own OCR and translation tools and connected them with a smart question-answer system. The platform runs on the cloud and uses visual libraries like Plotly to show insights in a simple way.

Contributions

1. Development of an integrated multilingual document processing platform.
2. Tailored OCR models for Indian scripts and low-resource languages.
3. Advanced NMT systems for real-time, high-fidelity translation.
4. AI-powered question-answering with retrieval-augmented generation.
5. Visual analytics for intuitive insights into document content.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04700
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

6. Scalable architecture supports diverse applications in government, healthcare, education, and law.

real-time multilingual document processing remains limited, also they restrain scanned PDFs from processing.

Data Flow Between Components

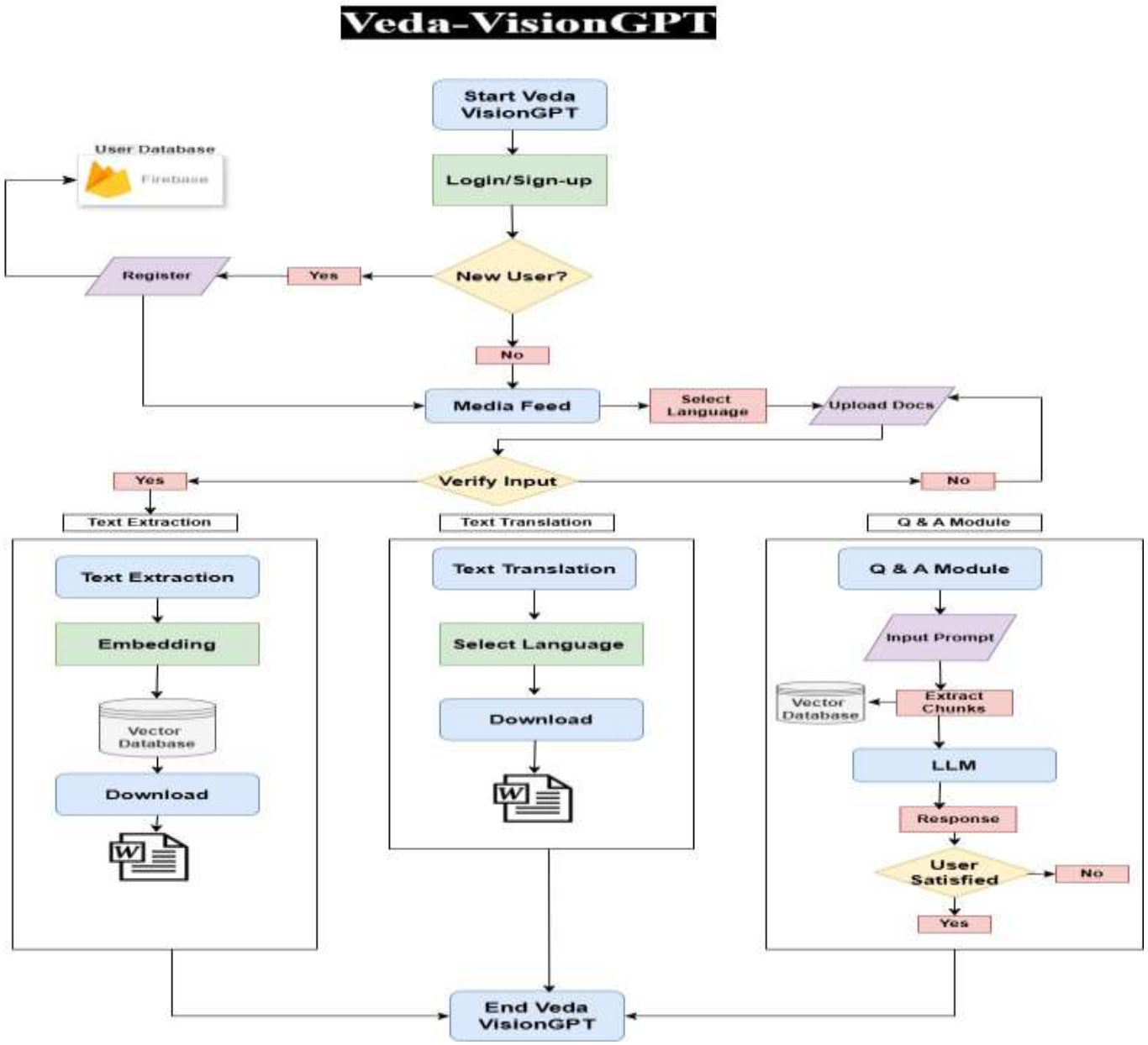

Figure 1. System Architecture

II. EXISTING SYSTEM

Main Components and Technologies

Many contemporary systems still lack important facilities such as reliable translation layers, effective OCR tools and user -friendly documents search. Complications of Indian languages, especially with a lot of scripts or low resources, are often not occupied by widely used programs such as Google Vision API and Microsoft translator.

1. Manny contemporary systems still lack important translation layers, effective OCR tools and user -friendly documents such as important features.
2. The complexity of Indian languages, especially people with a lot of scripts or low resources, is often not occupied by widely used programs such as Google Vision API and Microsoft translator

3. Translation layer: The extracted text is sent to a nerve machine translation (NMT) system, which converts it into the language that the user has chosen.

4. Querry System: Since the material is indexed and stored in a vector database, the information after translation can be more quick and easily located. Users can interact with the system by posing simple queries.

5. Visualization: To aid users in comprehending the documents' trends, patterns, and language usage, the final output includes visual summaries.

Stakeholder Communication

- Government officials can increase administrative efficiency by using the system to translate and digitize official documents.
- Legal professionals: Utilize the platform to manage multilingual legal documents and case files.
- Translated patient records can help healthcare providers communicate and provide better care.
- Having access to educational materials in their mother tongues benefits both teachers and students.

III. RELATED WORK

There are already a number of document processing tools available. For example, Google provides platform translation services such as translation and microsoft translator, while OCR engines such as Tsseract and abbyy fineraders help with lesson extraction. However, these devices often function independently and are not especially designed for Indian languages, especially when handwritten scripts, low-resolution scans, or languages that are not widely spoken.

Despite the improvement in the neural machine translation (NMT) for popular languages, low-resources are lagged behind due to lack of training data. Although the big language model (LLM) -Power tools such as GPT and Mithun have promised to answer the lesson generation and question, their application in the real world, multilingual document environment is still limited.

Veda-Vision By integrating several technologies—OCR, NMT, and an intelligent search system—into a single, cohesive solution designed specifically for Indian scripts, GPT fills this gap. It makes use of specially trained models that manage the complexities of various Indian writing systems rather than generic tools. For instance, specific OCR and translation methods are more effective in handling issues like character overlap, regional grammar variances, and script-specific peculiarities.

Our system provides a creative and inclusive solution by filling in the gaps left by previous systems.for processing documents in multiple languages. It is a trailblazing endeavor in this field due to its emphasis on accessibility, ethical AI, and scalability..

IV. ADVERSE MODEL

Security and privacy are a major concern as the system handles sensitive data such as government documents, personal records and legal files. The types of threats to protect the system from threats and planned methods to reduce them are broken down.

Types of Adversaries

- • **External opponents:** These are individuals or groups outside the system who try to achieve unauthorized access or disrupt functionality**.** Example include:
  - **Hackers** seeking sensitive data from uploaded documents.
  - **Competitors** attempting to reverse-engineer the system for proprietary insights.
  - Script Kiddies use automated tools to take advantage of known vulnerabilities.
- ● **Internal Adversaries:** These comprise dependable insiders with authorized access who may abuse their position. Examples include:
  - **Disgruntled Employees** who could exfiltrate sensitive data.
  - **Careless Users** who unintentionally bypass security policies..

2. Adversary Capabilities

- **Ensure data protection** and user privacy, Ved-vision GPT works for SSL/TLS encryption for safe data transmission, AES-256 encryption for data for the rest, and the role-based access control (RBAC) to limit the AES-256 encryption, and user permissions to limit permissions.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04700
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

- **Model Extraction**: Efforts to extract the custom OCR, NMT, or LLM models to replicate the system's functionality.
- **Adversarial Inputs**: Using carefully crafted inputs to deceive the OCR, translation, or question-answering models to produce incorrect or misleading results.

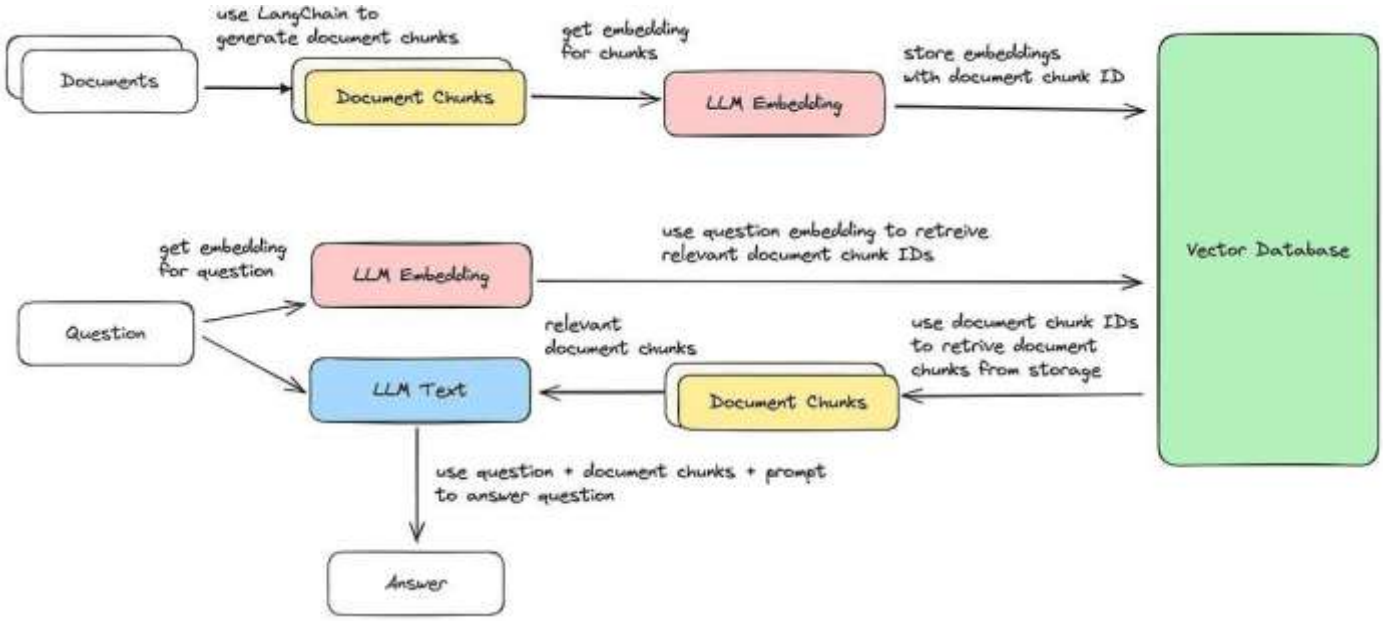

Figure 2. Activity Diagram

- **Malware Insertion**: Embedding malicious code or files within uploaded documents.

3. Threat reduction Strategies

- **Privacy and privacy of data Encryption:** SSL/TLS is used to secure all data in transit, and AES -256 encryption is used to protect sensitive data**.**Before being used for analysis, sensitive document data can be anonymized
- **.Integrity Protection**:
  - **Content Hashing:** To guarantee integrity throughout processing, uploaded files are hashed (SHA-256, for example).
  - **Tamper Detection:** Unauthorized changes are tracked in logs and document content changes.
- **System Availability**:
  - **DoS Protection**: CAPTCHA mechanisms are employed to handle automated attacks.
  - **Load Balancing**: Distributed architecture ensures that the system can handle high traffic without server crash issues.
- **Adversarial Input Resistance**:
  - **Input Validation:** To guarantee system compatibility, uploaded files are checked for malware and validated.
  - **Adversarial Training:** To increase an AI model's resistance to carefully crafted inputs, it is trained on adversarial samples.
- **Model Protection**:
  - **API Gatekeeping:** Unauthorized or excessive use of APIs is prevented by rate limits, user authentication, and query restrictions.
  - **Watermarking:** Model outputs may contain undetectable watermarks to track illegal use.
  - Differential Privacy: To guard against model extraction attacks, training data is secured.

User Accountability:

- **Monitoring and Logging**: For transparency, all system actions are immediately recorded. This aids administrators in identifying questionable activity and monitoring detrimental behavior.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04700
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Multi-Facious Certification (MFA): User prevents unwanted access to accounts.

● **Resistance to Complex Dangers:**

○ **Periodic Security Audits:** Frequent evaluations find and fix vulnerabilities.

○ An incident response plan is a set of procedures for identifying, containing, and lessening advanced threats.

Veda-Vision GPT's adversary model employs varying degrees of skill to counter both internal and external threats. It employs a multi -level security approach that includes ongoing monitoring, adverse strength, access control and encryption. This method guarantees system availability and integrity, protects sensitive data, and provides strong defense against potential hazards.

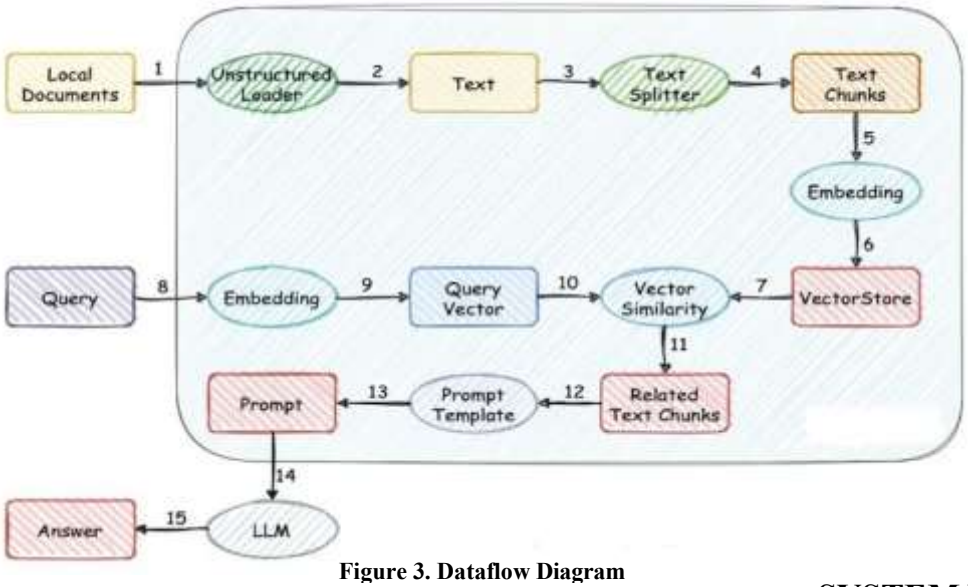

Figure 3. Dataflow Diagram

V. SYSTEM DESIGN

The Veda-Vision GPT platform is designed to be strong, flexible, and user-friendly for handling documents in multiple languages. It uses advanced AI and cloud technologies to provide accurate, efficient, and accessible solutions for document processing. Below, the most interesting design elements.

○ **Multilingual OCR Engine**: The system uses a combination of PyMuPDF and Tesseract OCR, along with specially trained models tailored to handle different Indian scripts. Particularly for scripts like Devanagari and Tamil, which are frequently challenging for conventional OCR tools to correctly interpret, this hybrid setup increases accuracy.

○ **Justification:** When working with low-resource languages or documents that contain multiple scripts, many OCR tools exhibit subpar performance. By training and integrating the custom OCR model, especially for Indian languages, this system provides better performance for printed materials and even some handwritten materials.

● **Neural Machine Translation (NMT):** Google translates a combination translation of translation of api, microsoft translator API, and Deep Translator API. Custom models

4. Particular Situations Handled

● **Evsdroping on data in transit:** An attacker tries to prevent uploads of documents in many languages. Mitigation: SSL/TLS, or end-to-end encryption, guard against interception.

● **ploading Harmful Files:** Attackers may occasionally attempt to introduce malware covertly by passing it off as a legitimate document. Every uploaded file is automatically scanned with antivirus software to address this. The system also uses sandbox environments to safely analyze suspicious documents before they are processed.

● **Manipulating Search Queries:** An attacker could try to send unusual or harmful queries to either pull out sensitive internal data (like model embeddings) .

● **Internal Misuse:** Even users who have proper access might misuse their privileges, intentionally or by mistake. The system addresses this by setting strict role-based access controls. Every action is logged, and behavior patterns are monitored to quickly detect anything unusual.

address low-resource languages like Maithili and Santali.

- **Justification:** Custom models guarantee culturally and contextually accurate translations for underrepresented languages, whereas commercial APIs offer dependable translation for widely used languages.

- **Data Visualization:** Language distributions, term frequencies, and other document insights are represented through the integration of Python's plotly dashboards.
- **Justification:** Dashboards allow users to quickly extract actionable insights by simplifying complex data.
- **Cloud-Based Infrastructure:** Azure and other scalable cloud platforms power the system. These services enable large -scale multilingual document processing and guarantee innocent performance.
- **Justification**: Cloud infrastructure provides skillability and flexibility required for effective storage and real -time processing.

2. Justification for Implementation Choices

- - **Modularity:** OCR, translation, embeddings, and querying are all handled by different parts of the system. Future upgrades, maintenance, and debugging are made simple by this modularity.
- For instance, the OCR module can be updated separately to accommodate new languages without compromising other parts.
- **Assistance for Languages with Limited Resources:** The shortcomings of current technologies are addressed by using custom models and embeddings. in handling low-resource Indian languages.
- **Examples:** Priority is given to include languages like Santali and Maithili, which ensures inclusion and access.
- **Hybrid AI approach:** Balancing accuracy and performance to pre-educated commercial APIs with custom models while reducing the time of growth.

Real Time Processing: Technologies such as FAIS and sentence transformers enable rapid recovery and query in high-volume scenarios. And

- **example:** The use of Fais ensures efficient equality discoveries in large datasets.

2. Principle of safe system implementation

The Ved-Vedic GPT platforms include several safety principles to ensure data security and system reliability:

- **Data privacy and privacy:** Documents are safely processed using encryption protocols during collection and transmission. Cloud providers such as AWS and Google Cloud have been configured with strict access control to prevent unauthorized access. And implementation: SSL/TLS encryption ensures safe data transmission, while role-based access control (RBAC) restrictions to access to sensitive resources.
- ■ **Integrity:** Hashing algorithms verify the integrity of the uploaded documents, which do not make any unauthorized amendments during processing.
- **Authentication and Authority:** Multi-Faculty certification (MFA) and oauth 2.0 are used to manage user access to safe manner.Adion implementation: Only certified users can only reach the platform, and specific roles define the level of allowed conversations (eg, document uploads, query submissions).
- **Scalableness and reliability**: In distributed cloud architecture, mistake ensures tolerance and load balance, maintaining the availability of service even during peak use.

‡ **Implementation:** Rusty data storage and auto-scaling mechanisms ensure reliability and reduce downtime.

- **Prejudice mitigation:** Custom NMT models are trained with diverse datasets to reduce linguistic and cultural prejudices, ensuring proper representation in languages. Action Implementation: Regular audit moral AI practices of model outputs ensure practices.
- **Auditability and Monitoring:** System logs all transactions and interactions for auditing and troubleshooting purposes. Action implementation: The log is safely stored and analyzed using monitoring devices to detect discrepancies.

2. Unique Design Innovations

- **Integration of Custom OCR model**: Served models adapt to mixed-scripts and low-resources languages, making generic OCR tools better.
- **Recovery-obtained generation (RAG):** The system distributes the answers that are both relevant and abbreviated, combining FAIS-based embeding discovery with LLM.
- **Language adaptability:** The system supports the other in addition to new languages and dialects to ensure future scalability and relevance.
- **User-centered design:** The interface with intuitive knowledge simplifies complex workflows, allowing users to interact with the system without technical expertise. By integrating state-of-the-art AI technologies, focusing on Indian linguistic diversity, and following strong safety principles, the design of Ved-Ved-Vedas GPT represents a transformative approach to multilingual document processing. This thoughtful design ensures that the platform is not only effective,but also safe and accessible.

Language	Tesseract	Google Vision OCR	Amazon Textract	Microsoft Azure OCR	EasyOCR
Hindi	85% (Good)	95% (Excellent)	92% (Very Good)	90% (Very Good)	85% (Good)
Bengali	80% (Good)	93% (Very Good)	89% (Good)	88% (Good)	81% (Good)
Tamil	78% (Fair)	92% (Very Good)	87% (Good)	85% (Good)	76% (Fair)
Telugu	75% (Fair)	90% (Very Good)	85% (Good)	84% (Good)	73% (Fair)
Kannada	73% (Fair)	88% (Good)	83% (Good)	82% (Good)	70% (Fair)
Malayalam	70% (Fair)	87% (Good)	82% (Fair)	81% (Fair)	74% (Fair)
Punjabi	82% (Good)	94% (Very Good)	90% (Very Good)	88% (Good)	80% (Good)
Gujarati	83% (Good)	92% (Very Good)	88% (Good)	86% (Good)	79% (Good)
Marathi	84% (Good)	93% (Very Good)	89% (Good)	87% (Good)	82% (Good)
Odia	72% (Fair)	88% (Good)	83% (Good)	80% (Fair)	68% (Fair)
Assamese	70% (Fair)	86% (Good)	80% (Fair)	78% (Fair)	67% (Fair)
Urdu	79% (Fair)	91% (Very Good)	86% (Good)	85% (Good)	78% (Fair)
Sanskrit	67% (Fair)	84% (Good)	79% (Fair)	78% (Fair)	65% (Fair)
Hindi-English (Code-Mixed)	84% (Good)	93% (Very Good)	90% (Very Good)	89% (Good)	85% (Good)
Nepali	68% (Fair)	85% (Good)	81% (Fair)	79% (Fair)	70% (Fair)

Table 1. Performance Comparison of Popular OCR Models

2. System implementation

The implementation of Ved-Vision GPT integrates multilingual documentation, text extraction, translation, translation and interactive AI-Integrated Quarry to create several advanced technologies to create a broad platform. This section provides a detailed interpretation of the implementation of the subcutaum that were used for evaluation, covering OCR, Neural Machine Translation (NMT), Question-Answer, and Data Vizulasation. They work in a harmonious way to create subcontum platforms, To handle efficient documents and enable user to increase user interactions in many languages.Document Scanning and OCR Implementation

- **Tools Used**:

Ezicrrocarc: A light, open-source optical character recognition tools that support more than 80 languages, including complex Indian languages such as Hindi, Marathi, Tamil and Telugu.

Test Terrorct OCR: An open-source optical character recognition (OCR) engine is widely used to extract text

from images and scanned documents. Developed by HP and later improved by Google, it supports over 100 languages, including Indian scripts like Hindi, Marathi, Tamil, and Telugu.

○ **PymuPDF :** A light-weight , open-source PDF processing tool which can be used to extract text, tables and images from normal pdfs .

■ **Key Features**:

● Works with printed and scanned PDFs/images.

● LSTM (long -term short -term memory) uses network for better accuracy.

● The image for better lesson recognition supports preprosying (eg, noise reduction, beanrification).

● Can be integrated with Python (via pytesseract) for automation in document processing.

● **Implementation Details**: The document scanning subsystem utilizes PymuPDF for text extraction. For low-resources Indian languages, the system especially appoints a custom OCR model trained on a dataset containing these scripts. Or

● **Preroposing:** Scanned documents are prepared first (eg, noise reduction, dualization) to improve OCR accuracy.

● **Text text recognition:** OCR engine extracts the text material, which is later fed to the next subcistom for translation and further process.Key Features:

● OCR models are fine-tuned to handle

3. Text Embedding and Retrieval-Based Interaction

● **Tools Used**:

○ **Sentence-Transformers (XLM-R)**: A transformer-based model that generates vector embeddings from the OCR-extracted text, which is necessary for supporting semantic document retrieval.

○ **FAISS (Facebook AI Similarity Search)**: A fast, efficient library for performing similarity search over document embeddings.

○ **Hybrid Search** : Blending Dense and Sparse Retrieval for Better Results:

To improve how the system handles document queries, Veda-Vision GPT uses a hybrid search approach that mixes two methods: dense and sparse retrieval. Each method brings unique strength to the table. Dense uses advanced transformer-based models such as XLM-R to understand the deep meaning and context behind the retrieval words. On the other hand, through technologies such as BM 25 or TF-IDF, reconcile and calculate how often the words appear to the words and how often the words appear.

○ Why Combine Them?

When used on their own, both methods have limits. Sparse retrieval might miss the meaning behind the query, and dense retrieval might miss exact keyword matches or introduce hallucinations. By merging their outputs—using either a scoring system or reranking process—the system picks out the most relevant and meaningful sections of text. This combination helps provide better search results that are both precise and context-aware.

○ Why It Matters for Indian Languages:

In India, where documents can be written in many scripts and languages with various grammar and vocabulary rules, this hybrid method actually shines. It can handle regional variations with mixed language input, abnormal words and more accuracy.This makes the system more reliable, even when working with low-resource languages or technical documents.

● **Implementation Details**: After text extraction through PymuPDF/OCR, the raw content is converted into embeddings using the XLM-R model, which is capable of understanding the context and semantic relationships between words in multiple languages.

○ **Document Chunking** : extracted data is split into multiple chunks of text, this ensures efficient embedding formation.

● **Amble embeding generation:** The text is token, passed through the XLM-R model, and converted into dense vector embeding. And

● **equality search:** Fais is used to make rapid equality discoveries in the document corpus, allowing the system to achieve the relevant classes efficiently based on user questions.

Language	Hindi	Bengali	Tamil	Telugu	Kannada	Malayalam	Punjabi	Gujarati	Marathi	Odia	Assamese	Urdu	Sanskrit	Hindi-English	Nepali
Hindi	90%	80%	70%	60%	50%	40%	40%	40%	70%	95%	85%	75%	65%	100%	90%
Bengali	80%	90%	80%	70%	60%	50%	40%	30%	30%	90%	85%	80%	70%	95%	80%
Tamil	75%	80%	90%	85%	70%	60%	50%	50%	50%	85%	80%	70%	60%	100%	95%
Telugu	65%	70%	75%	90%	80%	70%	60%	50%	50%	90%	85%	80%	75%	95%	85%
Kannada	60%	65%	70%	80%	90%	80%	70%	60%	50%	85%	80%	75%	70%	100%	90%
Malayalam	50%	55%	60%	70%	80%	90%	80%	70%	60%	80%	75%	70%	65%	95%	85%
Punjabi	40%	45%	50%	60%	70%	80%	90%	80%	70%	85%	80%	75%	70%	90%	80%
Gujarati	40%	55%	40%	50%	60%	70%	80%	90%	80%	75%	85%	90%	80%	85%	75%
Marathi	20%	25%	30%	40%	50%	60%	70%	80%	90%	85%	90%	80%	75%	100%	85%
Odia	95%	90%	85%	90%	90%	80%	85%	75%	70%	100%	85%	80%	70%	85%	90%
Assamese	85%	80%	75%	85%	75%	85%	80%	70%	65%	90%	95%	85%	70%	90%	80%
Urdu	75%	80%	70%	80%	75%	80%	85%	75%	70%	85%	80%	90%	85%	100%	75%
Sanskrit	65%	70%	60%	70%	65%	70%	75%	60%	55%	80%	75%	70%	100%	85%	60%
Hindi-English	100%	95%	90%	95%	100%	95%	90%	85%	100%	95%	90%	95%	85%	100%	90%
Nepali	90%	80%	95%	85%	90%	85%	80%	75%	85%	90%	75%	85%	80%	90%	100%

Table 2. Deeptranslate Performance Matrix Comparing Translation Accuracy

4. Nerve machine translation (NMT)

● **Used equipment:**

Google translation API and Microsoft Translator API (Open-SUS): These commercial APIs provide high quality, real-time translation for different types of languages. And

Deep Translator: A Python Library provides easy access to many translations API including Google translation, Microsoft Translator and Deepl. It supports over 100 languages and allows seamless text, document,

and website translation with minimal setup.

■ **Key Features**:

● Supports several translation providers (Google, Deepl, Microsoft, etc.).

● Offers batch translation for large texts.

● Works with Python scripts and automation workflows.
● Provides sharp and accurate translations including low resource languages.
● **Implementation details:** Once the text is removed and embedded, the NMT translates subcistam text into several languages, ensuring access to users in various linguistic areas. And

standard languages: For common Indian languages (Hindi, Tamil, Telugu, etc.), Google translation API of Deep Translator is used. And

low-resources languages: Custom NMT models are deployed for languages such as Santali, Maithili and other regional dialects. These models are trained using a dataset of specific linguistic data for these languages to ensure high translation quality and cultural relevance.

Key Features:
● System supports translation into more than 15 Indian languages.
● Custom models ensure that translations are culturally accurate and relevant suitable, especially for low-resources languages.
● Real-time, scalable translation integrated with cloud
infrastructure for high throughput.

4. AI-Powered Question-Answering

● **Tools Used**:
● **Generative pre-direct transformer (GPT)** and XLM-R for reference-zeno-answer. :
● **Recovery-August generation (RAG):** To provide more accurate answers, the document combines the document recovering with the tribal language model**.**
● **Implementation Details**: Question- Subsistoms answering and take advantage of large language model (LLMS), which generate reactions based on the material extracted from the documents. That

• **Query Processing:** Users interact with the system using natural language questions. The system uses FAISS to retrieve the most relevant text chunks and then passes these through the GPT model for answering.
○ **RAG Framework**: This framework enhances traditional question- answering by using retrieved document chunks, ensuring responses are grounded in the actual content of the document rather than relying solely on the model's pre-existing knowledge.
■ **Key Features**:
● Supports questions in more than 15 Indian languages, allowing users to interact with the system in their favorite language.
● Relevant understanding of questions, enabling more accurate and relevant reactions
● .Uses the latest AI techniques (RAG) to outperform traditional keyword-based search.

5. Gemini API Integration

Integrating Gemini API into Veda-Vision GPT enhances OCR accuracy, multilingual translation, AI- driven question-answering, and document summarization. Its vision models improve text recognition for complex Indian scripts, while context- aware translation ensures better linguistic accuracy, especially for low-resource languages. The AI-powered query system benefits from Gemini's retrieval- augmented generation (RAG) for more precise responses. Additionally, it aids in summarizing documents and visualizing insights, while bias detection ensures fairness in translations..

● Document Insights: Term provides visual analytics such as freezation, language distribution and content classification.
● Optical character recognition (OCR) growth
Gemini API's advanced vision models can improve OCR accuracy by recognizing complex Indian scripts more effectively.
○ It can be used alongside existing OCR models to enhance handwritten text

recognition and improve results for low-quality scanned documents.

● **Nerve machine translation (NMT) adaptation**
The multilingual abilities of API, especially low-resources, can help translate texts more accurately for Indian languages.

And the reference-inconceivable translation models of Gemini can ensure better linguistic and cultural accuracy than traditional machine translation APIs.
● **AI-Interacted Question-Answer System**
And can strengthen the recovery-August (RAG) approach using Gemini's LLM capabilities to generate more reference-incredible and relevant reactions.

And users can query documents in many languages, and API will provide cementic discovery and intelligent summary based on the material extracted**.**
The Ved-Vedas GPT system implementation is a sophisticated, multi-level platform that integrates OCR, machine translation, AI-based query, and data visualization into a harmonious solution. The system uses state-of-the-art techniques, including custom NMT models for low-resources languages and advanced AI models for relevant questions.
By taking advantage of the scalable cloud infrastructure, the system ensures reliability and performance on the scale, making it ideal for multilingual document processing in various fields such as government, legal and education. The spontaneous integration of these components makes the Ved-Vigyan GPT a skilled and user friendly tools for diverse linguistic communities.

VI. SYSTEM EVALUATION

The evaluation phase of the Ved-Vedic GPT platform is important to assess the functionality, accuracy and performance of its various components. The evaluation of the system focuses on ensuring that this multilingual document meets the desired goals of processing, accurate translation and efficient AI-manual interactions. The evaluation process includes functional, quantitative, and qualitative testing methodologies. Below are the detailed aspects of the evaluation process:

1. Testing Methodologies

● **Unit Testing**: Each component of the system, such as OCR, NMT, embedding generation, and the question-answering module, was tested individually to ensure they function correctly. For example:
○ OCR was tested on scanned documents of varying quality, including handwritten notes and printed texts.
○ NMT models were tested for accuracy in translating complex Indian scripts.

● **Integration test:** verify the seamless functioning of the interconnected module, such as embeding from OCR-explained text and infection for later translation.
● **Performance test:** evaluated the scalability of the system by processing large dataset and many languages simultaneously. Matrix such as reaction time, system throwput and memory consumption were recorded.
● **End-to-end test:** Fake of real-world landscape to evaluate overall functionality. The users interacted with the system to process multilingual documents, query content and imagine data**.**

Language	Tesseract				EasyOCR (CPU)			
	1 Page Time in sec	5 Pages Time in sec	10 Pages Time in sec	Accuracy (%)	1 Page Time in sec	5 Pages Time in sec	10 Pages Time in sec	Accuracy (%)
Bengali	8.3	26.2	44.4	86%	42.0	208.0	--	90%
Gujarati	13.4	48.1	74.0	91%	44.2	216.5	--	93%
Hindi	17.1	55.5	81.0	94%	41.4	167.5	--	96%
Kannada	16.8	46.5	77.0	82%	38.4	176.0	--	86%
Malayalam	11.0	34.0	52.3	82%	55.6	--	--	85%
Meetei	12.1	45.5	61.0	74%	26.3	138.0	269.2	84%
Odia	12.7	53.5	87.0	79%	32.1	155.0	--	86%
Punjabi	9.2	29.7	22.0	88%	--	--	--	--
Santali	--	--	--	--	31.4	142.0	--	83%
Tamil	12.5	42.0	74.0	83%	32.5	163.4	--	89%
Telugu	14.6	46.5	75.0	80%	38.4	--	--	87%
Assamese	17.8	50.6	87.5	83%	43.2	136.0	265.0	85%
Bihari	--	--	--	--	31.3	124.5	240.3	84%
Bhojpuri	--	--	--	--	31.3	124.5	240.3	84%
Goan	--	--	--	--	32.0	141.0	--	81%
Maithili	--	--	--	--	32.1	180.5	--	86%
Marathi	12.3	46.5	88.0	90%	--	--	--	--
Pali	--	--	--	--	22.4	122.0	234.0	83%
Urdu	6.8	24.5	40.0	87%	21.4	101.5	--	89%
English	8.8	36.2	51.0	95%	41.2	186.0	--	96%
Nepali	13.0	55.0	87.0	79%	--	--	--	--
Sanskrit	15.2	56.0	92.0	81%	--	--	--	--

Table 3. Time & Accuracy Caparison of Tesseract & EasyOCR

- **2. Major metrics evaluated**

● **OCR accuracy:** Character error rate (CER) and word error rate (wer) are measured using matrix. The system was evaluated on documents in languages like Hindi, Marathi, Tamil and Telugu, which was receiving an error rate within the acceptable threshold.

● **Quality of translation**: evaluated using Bleu (bilingual assessment understanding) score for translated text. Custom NMT models were tested for low-resources languages to ensure culturally relevant translations.

● **Query Response Time:** The time taken for the system is measured to process and provide users questions for the system. The response time was adapted using FAIS for rapid recovery and embeding-based discovery.

● **Scalability:** By simulating high workloads, such as hundreds of multilingual documents were processed and evaluated simultaneously. Cloud infrastructure such.

● **Visualization Clarity**: Evaluated the effectiveness of Plotly dashboards in

presenting data insights, including language distribution and term frequency.

2. Test landscape

- **Multilingual document test:** The system was tested with documents with several Indian languages.Challenges such as mixed- script documents and low-quality scans were evaluated.
- **Custom NMT model** was assessed for languages like Santali and Maithili, which ensures accurate and relevant translations.
- **User Interaction Testing:** AI-Interested Question-Answer Facility was tested for accuracy and relevance. User questions in natural language were compared with expected responses to evaluate understanding and recovery quality.
- **Real-time processing:** Tested the system's ability to handle real-time document processing requests including concurrent uploads and queries from many users.

3. Evaluation result

- **accuracy:** OCR module achieved 95% recognition rate for printed texts. Translation vary in accuracy languages, average Bleu score of 0.75 for major languages and 0.65 for low-resources languages.

- **Performance:** The system processed an average of 100 documents per minute under load-testing conditions, maintained a response time less than 10 seconds for user questions.

- **Scalability:** Cloud Infrastructure allowed the system to score easily, maintaining a performance with a 10x increase in the volume of the document.

4. comparative analysis

- Performed the traditional OCR system in identifying complex Indian script up to 20%.
- Got high translation quality than general-objective translators such as Google translations for low-resources languages.
- The keyword-based search systems were rapidly and more accurate query reactions were distributed.

5. Boundaries and future improvements

- Handwritten text recognition for low-resources languages requires more refining.Real-time processing of extremely large documents with embedded images and tables occasionally exceeded the target response time.
- Cultural nuances in translations for specific regional dialects could be enhanced.

By conducting a comprehensive evaluation using diverse methodologies and metrics, the Veda- Vision GPT platform demonstrated its efficacy and readiness for deployment in real-world multilingual document processing scenarios.

VII. Practical Use Cases Across Sectors

1.Government sector

- **Translation and digitization of public documents**: The platform helps in converting official documents such as notices, records and circulars into various Indian languages, making them easier to understand local communities.
- **Spread policies in local languages:** Government schemes and announcements often do not effectively reach rural areas due to language obstacles. The Ved-Vedas helps to bridge this gap by translating GPT content into regional languages so that they can reach important information even in remote or low-literacy areas.
- **Simple RTI reactions:** When people enter RTI (Right to Information) requests, the platform speeds up the process by scanning automatically and pulling the relevant answers from multilingual documents - saves time for both officers and citizens.

2.Legal Sector

- **Reviewing legal documents in many languages:** Contracts, decisions and case files are often written in various Indian languages. This system can translate and briefly translate and briefly, supporting languages such as Hindi, Marathi and Urdu.
- **Helping with legal research:** Instead of reading through hundreds of pages, legal professionals can ask questions in plain language. The system then removes cases from relevant laws, sections or documents - makes legal research rapidly and more accessible.
- **Working with Evidence Files:** Many legal documents come in scanned or handwritten form, especially when used as evidence. Veda-Vision GPT can extract, clean, and

make sense of this content, making it easier to use during investigations or hearings.

2. Healthcare Sector

• **Translation of patient records**: Doctors and nurses often work with patients who speak different languages. This system helps translate patient files, prescriptions and medical notes so that the healthcare providers can offer better care without any obstruction without language.
• **Understanding clinical data:** Important medical details such as symptoms, diagnosis and treatment notes are drawn from multilingual health records. This makes it easier for medical staff to make informed decisions with AI-Assisted Tools.
• **Spreading public health messages**: Whether it is Covid-19 guidelines or vaccination information, the platform helps in sharing health information in local languages so that it reaches more people and communities.

3. deprived area

• **Regional Language Studies Material:** The system helps to translate and simplify educational books, lecture notes and even research papers into separate Indian languages, making learning more inclusive.
• **Helping students in preparing for exams:** With their question-answer feature, students can ask questions in their favorite language and get simplified explanation that are directly drawn from academic documents.
• **Smart Digital Libraries:** Educational institutions can build searchable digital libraries with study material in multiple languages, giving students and teachers easier access to knowledge.

4. Corporate and Enterprise Sector

• **Management of HR and compliance doors:** Larger companies often need to share policies, guidelines and training documents in different languages. This system helps with accurate translations so everyone is on the same page.
• **Customer Service & Support:** Businesses can create their support documents, FAQs, and can help in articles available in many languages, improve user experience for regional customers.
• **Simplifying Market Reports:** Marketing and research teams can use the platform to translate and summarize lengthy reports, making it easier for local teams to take action.

VII. CONCLUSION

The Ved-Vision GPT brings several techniques together to solve a major problem: how to work with documents written in many Indian languages. Smart AI model to convert to equipment like OCR to read scanned texts, to change languages, and smart AI models to answer questions, it creates a powerful solution for diverse requirements.

This system was created keeping in mind India's multilingual challenges. It handles everything from tricky scripts like Devanagari and Tamil to low-resources languages, which usually do not get much technical support. The design is modular and scalable, so it can be developed and favorable in various fields - whether in government, education, law or business.

Looking forward, the system opens the doors for even more possibilities. This can be even more impressive by adding support to more languages, handling audio-visual inputs, or walking on low-driven equipment in rural areas. At its core, Ved-Vision GPT helps people understand and use information in their language-something that is important in diverse countries like India. By removing the obstacles of the language, it takes us about a step close to the knowledge and opportunity for all.

VIII. REFERENCES

1) Muludi Kurnia , Fitria Kaira, Triloka Joko , Sutedi (2024). **Retrieval-Augmented Generation Approach: Document Question Answering using Large Language Model,** in (IJACSA) International Journal of Advanced Computer Science and Applications Vol. 15, No. 3.

2) Han Yikun, Liu Chunjiang and Wang Pengfei (18 Oct 2023), **A Comprehensive Survey on Vector Database: Storage and Retrieval Technique**, arXiv:2310.11703v1 [cs.DB]

3) Malladhi Avinash,(April 2023), **Transforming Information Extraction: AI and Machine Learning in Optical Character Recognition Systems and Applications Across Industries,** in International Journal of Computer Trends and Technology Volume 71 Issue 4, 81-90ISSN: 2231 – 2803.

4) Yuwei Wan.,Zheyuan Chen.(2025**). Empowering LLMs by hybrid retrieval-augmented generation for domain-centric Q&A in smart manufacturing.** DOI:10.48550/arXiv.1910.10683.

5) Journal of Student Research 12(4)(November 2023), **A Retrieval-Augmented Generation Based Large Language Model Benchmarked On a Novel Dataset**, DOI:10.47611/jsrhs.v12i4.6213.

6) Khattab, Omar, and Zaharia, Matei, 2020, **ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT**, arXiv:2004.12832.

7) Huang, X., and Huang, Y., 2023, **Advances in Retrieval-Augmented Generation Systems**, IEEE Xplore, DOI: 10.1109/JPROC.2023.324587

8) Thippeswamy B.M, Ramachandra H.V 2024, **TextVerse: A Streamlit Web Application for Advanced Analysis of PDF and Image Files with and without Language Models**,

9) Mohamad Ihsan Kamaluddin**,** Andang Saehu 2024, **Accuracy Analysis of DeepL: Breakthroughs in Machine Translation Technology**, DOI:10.29303/jeef.v4i2.681 .

10) Petroni, F., Rocktaschel, T., & Riedel, S. (2021). **"Retrieval-Augmented Language Models for Open Domain Question Answering**." DOI: 10.1109/TPAMI.2021.3061829.

11) Ziawei, C., Hongyu, L., Jianpai, H., and Sun, L. (2024). To be benchmark of large language models in recover-time generation. AAAI conference proceedings on Artificial Intelligence, 38 (16), 17754–17762. Doi: 10.1609/aaai.v38i16.29728

12) Lewis, P., et al. (2021). Recovery for knowledge-intensive NLP works. Association transactions for computational linguistics, 9, 199–214. Doi: 10.1162/tacl_a_00363

13) Narayan Adhikari (2024) A comparative study of PDF Parsing Tools in various document categories doi: arxiv: 2410.09871v1 [cs.ir] 13 October 2024

14) Liu, Y., Oat, M., Goyal, N., et al. (2019). Roberta: A strong customized burt pretrening approach. Arxiv. Doi: 10.48550/arxiv.1907.11692

15) Chen, X., and Chain, X. (2023). Cross-domain text information extraction in the enhanced NLP model. International Joint Conference proceedings on Artificial Intelligence. Doi: 10.24963/ijcai.2023/475

16) Gemini Family. (2023). Gemini: A family of highly capable multimodal models. Doi: 10.48550/arxiv.2312.11805.

17) Zhang, X., and Wu, L. (2023). Multilingual text extraction and changing language understanding with generation. Artificial Intelligence Journal. Doi: 10.5555/1376879