IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 9 ISSUE : 05 Print / Issue Publication Date: 08-Dec-2024

ISSN : 2455-2143

DOI : 10.33564/IJEAST.2024.v09i05.015

Indexed In

Google SCHOLAR OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 05, ISSN No. 2455-2143, Pages 129-134
Published Online September 2024 in IJEAST (http://www.ijeast.com)

VEDA-VISION GPT-AN AI-POWERED MULTILINGUAL DOCUMENT PROCESSING AND INTERACTION PLATFORM

Kunal M.Sapkal, Vinayak R.Gharge, Sanika J.Kadam, Rutuja N.Mulik, Prof. Vaibhav U. Bhosale
Dept. of Computer Science and Engineering
Karmaveer Bhaurao Patil College of Engineering,
Satara Maharashtra, India

Abstract:-**VEDA-VISIONGPT innovates multilingual document handling by unifying text recognition, language translation, and retrieval-augmented generation with interactive AI. This ground breaking platform extracts content from all types of documents in diverse Indian language sources, renders translations across more than 15 native tongues, and facilitates natural language queries. Employing advanced OCR and AI technologies, it offers comprehensive multilingual document management. Enhanced text analytics enable clear, logical information extraction. Aimed at government, legal, and academic sectors requiring precise language processing, VEDA-VISIONGPT revolutionizes cross-lingual information access and comprehension, transforming how diverse linguistic communities engage with content.**

Keywords-**Multilingual Document Processing, Optical Character Recognition (OCR) Neural Machine Translation, Conversational AI, retrieval-augmented generation (RAG).**

I. INTRODUCTION

The modern age demands an innovative platform for processing non-machine readable, multilingual documents, a challenge the Veda-Vision GPT addresses with remarkable ingenuity. This comprehensive platform integrates cutting edge technologies in text extraction, language translation, and AI-driven comprehension, specifically tailored for India's linguistically diverse landscape.

At its core, Veda-Vision GPT utilizes advanced Optical Character Recognition (OCR) algorithms optimized for Indic scripts, ensuring accurate text extraction from complex documents. The system's Neural Machine Translation (NMT) component then renders this text into 15 different Indian languages with high fidelity. Perhaps most innovatively, the platform incorporates large language models to enable interactive document exploration through natural language queries.

This unified solution represents a significant advancement over previous systems, which often struggled with the intricacies of Indian languages and lacked sophisticated querying capabilities. By combining OCR, NMT, and AI powered question-answering within a user-friendly interface, Veda-Vision GPT promises to revolutionize document processing across various sectors, from government administration to academic research. This survey paper examines the technical architecture, implementation challenges, and potential applications of Veda-Vision GPT. It explores how this transformative tool addresses the pressing need for efficient multilingual document handling in our increasingly globalized world, potentially reshaping how we interact with and derive value from diverse linguistic content. The system's question-answering module leverages retrieval augmented generation, combining GPT and BERT models to provide accurate, context-aware responses to multilingual queries.

This approach significantly out performs traditional keyword-based search methods, offering deeper understanding and more relevant results. Veda-Vision GPT also integrates advanced visual analytics tools, enabling users to intuitively grasp complex multilingual information through graphical representations of document content and language distribution.

The project's influence goes beyond simple translation; it promotes knowledge sharing and intercultural understanding. Its prospective uses in the legal, academic, business, and government sectors promise to remove language barriers and promote international cooperation. Veda-Vision GPT is a major advancement in how individuals and organizations deal with linguistic diversity in the digital age by tackling the challenges of multilingual document processing in a single platform, opening the door for more effective and inclusive information processing.

Existing Systems and Challenges Existing document processing systems, such as those using Optical Character Recognition and Neural Machine Translation, often struggle with the accuracy in processing complex Indian scripts like Hindi, Kannada, Telugu. Many current systems lack support for low-resource languages which result in poor translation quality. Additionally, handling large-scale document processing and efficient real-time interaction is challenging

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 05, ISSN No. 2455-2143, Pages 129-134
Published Online September 2024 in IJEAST (http://www.ijeast.com)

due to scalability limitations.
Overcoming Challenges: Veda-Vision GPT overcomes these issues by integrating customized OCR models for complex scripts, developing tailored NMT models for low-resource languages, and employing cloud based architecture for scalability, ensuring efficient real-time processing and enhanced multilingual support.

II. BACKGROUND & TERMINOLOGY

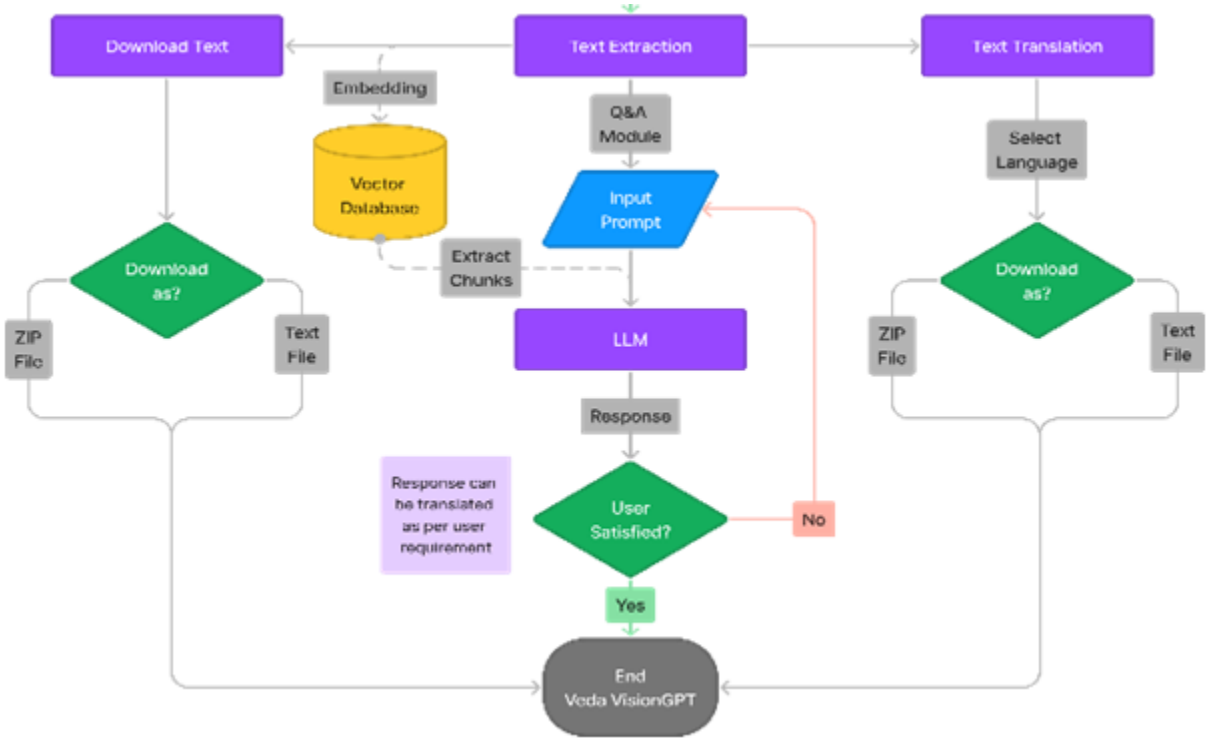
The Veda-Vision GPT addresses the challenges of non machine readable document processing, with a focus on Indian languages. By incorporating advanced technologies such as Optical Character Recognition, neural machine translation, AI based question-answering, it creates a highly efficient and user friendly platform. This system allows enhanced interaction with documents of all formats in multiple languages, ensuring efficiency, accuracy, and accessibility across various sectors, including government, legal, and academia.
OCR has played a crucial role in history of digital printed text since the 1950s, it was initially focused on English. These systems have evolved with time to support a broader range of languages, including Indian languages like Hindi, Marathi, etc. With the introduction of machine learning algorithms and deep neural networks, both extraction and translation technologies have significantly evolved in terms of accuracy and efficiency.
Veda-Vision GPT integrates these systems with AI-powered question-answering, enabling users to interact with multilingual documents through natural language prompts.
Key terms include OCR, which digitizes scanned text; NMT, which enables accurate translation using neural networks; and LLMs, which facilitate AI-powered question-answering. The project supports more than 15 Indian languages, enhances document accessibility, and emphasizes ethical AI practices to reduce bias while ensuring scalability.

Workflow of Veda-Vision GPT:

- Document Scanning: The system uses OCR to scan and convert physical or digital documents into editable and searchable text
- Text Extraction: OCR technology is applied to extract multilingual text from scanned documents.
- Embeddings: The X-Roberta embeddings model is used to convert the text into embeddings for efficient search for natural language query.
- Translation: NMT is used to translate the extracted text into 15+ Indian languages.
- AI-Powered Querying: Users interact with the system by asking natural language questions about document content, receiving precise answers via LLMs.
- Data Visualization: The system generates visual representations of the document's content and language distribution using tools like Power BI for quick analysis.

Workflow of Veda-Vision GPT

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 05, ISSN No. 2455-2143, Pages 129-134
Published Online September 2024 in IJEAST (http://www.ijeast.com)



III. TECHNOLOGY STACK

The Veda-Vision GPT project aims to provide a comprehensive multilingual document processing system using advanced technologies the step-by-step methodology outlines the system's design and development, detailing the technology stack used in the project.

1. Data Collection and Preprocessing: Gather a diverse set of multilingual documents that include Indian languages like Hindi, Marathi, Tamil, Telugu, and more.

Collect handwritten, printed, and digital documents to ensure the OCR system can handle multiple formats. Preprocess the data by converting all documents into a standard format such as images or PDFs, and label the documents according to language and content type for training purposes.

2. Optical Character Recognition:

Easy OCR: It is a broadly used and open-source tool for text extraction from images or non-machine readable documents. It supports a large array of languages, including complex Indian languages such as Marathi, Hindi, Tamil and Telugu. It is flexible and provides customization options, making it the best to use for multilingual document processing.

Google Vision API: It is a cloud-based OCR solution that offers enhanced accuracy for text extraction from printed and handwritten documents. It is highly scalable with a support of large array of diverse languages and enables real-time document processing in high-volume scenarios.

Script-Specific OCR Models: To ensure precise text recognition for Indian languages with unique character sets, customized OCR models are implemented.

3. Natural Language Processing:

Embeddings: The extracted content is converted into contextual vector embeddings using sentence-transformers (XLM-R), which is an encoder only model.

4. Neural Machine Translation:

Google Translate API: This API is integrated to provide real-time translation of the extracted text into 15 above native Indian languages, ensuring accessibility across diverse linguistic groups.

Microsoft Translator API: An alternative translation service offering similar capabilities, but with enhanced scalability for handling larger datasets in cloud environments.

Custom NMT Models: Custom translation models are built for low-resource languages such as Maithili and Santali, ensuring accuracy and cultural relevance for these underrepresented languages

5. AI-Based Question-Answering:

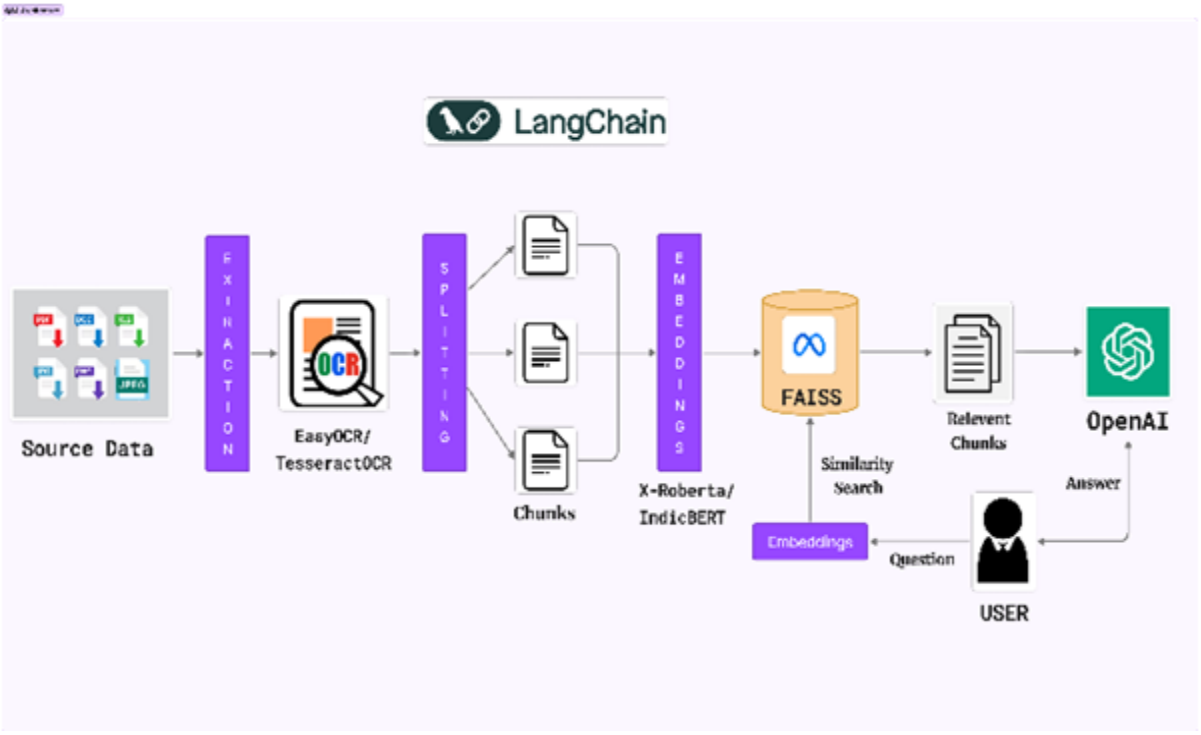
Generative Pre-trained Transformers: It is used to build an AI powered question-answering system that allows users to ask questions in 15+ languages and get relevant responses based on the document content.

XLM-R : XLM-R also knowns as sentence transformers is employed for vector embeddings, text chunking and retrieval tasks . These require deeper contextual understanding, particularly for answering specific questions

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 05, ISSN No. 2455-2143, Pages 129-134
Published Online September 2024 in IJEAST (http://www.ijeast.com)

from sections of a document.

Retrieval Augmented Generation



6. Vector Databases for Document Retrieval:
FAISS (Facebook AI Similarity Search): A powerful library used for similarity search and clustering of document embeddings. FAISS allows the system to perform fast, efficient searches across large document datasets.

7. Data Visualization:
Power BI:
Integrated to offer visual representation of document content, such as language distribution and term frequency, providing insights into document data in an easy- to understand graphical format.

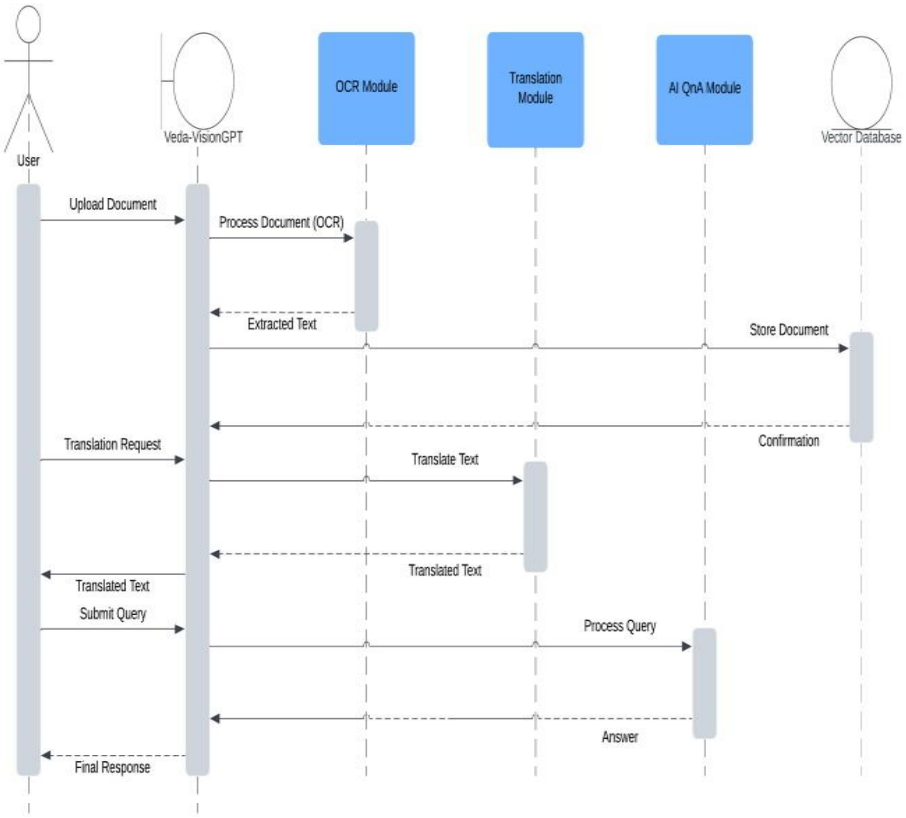
8. Cloud Infrastructure:
AWS/Google Cloud: The system leverages cloud platforms like AWS or Google Cloud to handle document processing and data storage. These cloud services ensure scalability, allowing the system to process large volumes of data while maintaining performance and efficiency.
This technology stack enables the Veda-Vision GPT project to offer an efficient, scalable, and highly responsive solution for multilingual document processing across various industries.

IV. PROJECT OBJECTIVES

1. Develop a one and only multilingual document processing platform which focuses on Indian languages, including complex scripts like Marathi, Tamil and Telugu.
2. Improve Optical Character Recognition accuracy for handwritten and scanned documents in multiple languages.
3. Integrate advanced Neural Machine Translation (NMT) models to support real-time translation across 15 Indian languages.
4. Implement an intelligent AI-powered question answering systems for seamless interaction with the documents using natural language queries.
5. Ensure scalability and efficiency through a cloud based architecture for handling large volumes of documents and real-time processing requests.

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 05, ISSN No. 2455-2143, Pages 129-134
Published Online September 2024 in IJEAST (http://www.ijeast.com)

Sequence Diagram of Veda Vision GPT



V. FUTURE RESEARCH DIRECTIONS

1. Enhance Optical Character Recognition capabilities by using advanced vision API's to improve accuracy with complex scripts and handwritten text in multiple languages.
2. Apply high level Neural Machine Translation (NMT) tools to allow access to other languages with low-resources.
3. Explore multimodal processing techniques that integrate text with images and graphs for an all inclusive understanding of documents.
4. Investigate bias mitigation strategies in AI models to ensure fair representation and accuracy across diverse linguistic contexts.
5. Optimize scalability solutions, including edge computing and distributed architectures, to support real-time document processing in resource - constrained environments.

VI. APPLICATIONS OF VEDA-VISION GPT

1. Government Services: Streamline administrative tasks by digitizing and translating official documents into multiple languages, enhancing public service accessibility.
2. Legal Sector: Aid legal professionals in managing multilingual legal documents, improving efficiency in case handling and legal research.
3. Healthcare: Facilitate the digitization and translation of patient records and medical documents, ensuring better patient care and communication.
4. Education: Provide students and educators with access to learning materials in multiple languages, promoting inclusivity in diverse classrooms.
5. Corporate Environment: Enable businesses to automate the translation of documents and communications, improving collaboration across linguistic boundaries.

VII. CONCLUSION

In summary, the Veda-Vision GPT project successfully addresses the intricacies of multilingual document processing by combining advanced technologies, including Optical Character Recognition, machine translation, and AI-powered question-answering systems. By emphasizing support for Indian languages and their unique scripts, the project enhances accessibility and communication across various sectors. Looking ahead, upcoming research will

focus mainly on overcoming existing challenges, refining the system's capabilities, and expanding its applications, thereby ensuring its effectiveness in bridging language barriers in diverse and dynamic environments.

VIII. ACKNOWLEDGMENT

We would like to express our sincere gratitude to everyone who contributed to the Veda-Vision GPT. Special thanks to our mentors for their precious guidance and support. We also appreciate the resources provided by our institute and the encouragement from our mates and family throughout this journey, which made this possible

IX. 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] Raffel, C., et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research, 21, 1-67. 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] Lecture Notes in Computer Science 3406:539-5473406:539-54,(February 2005), A Machine Learning Approach to Information Extraction, DOI:10.1007/978-3-540-30586.

[7] Budler Leona, Gosak Lucija and Stiglic Gregor(January 2023) Review of artificial intelligence-based question-answering systems in healthcare, DOI:10.1002/widm.1487

[8] Khattab, Omar, and Zaharia, Matei, 2020, ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT, arXiv:2004.12832.

[9] Huang, X., and Huang, Y., 2023, Advances in Retrieval-Augmented Generation Systems, IEEE Xplore, DOI: 10.1109/JPROC.2023.324587

[10] Borgeaud, Sebastian, et al., 2022, Improving Language Models by Retrieving from Trillions of Tokens, arXiv:2112.04426.

[11] Roberts, Adam, et al., 2020, Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, JMLR 2020, Vol. 21

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

[13] Jiawei, C., Hongyu, L., Xianpei, H., & Sun, L. (2024). Benchmarking Large Language Models in Retrieval-Augmented Generation. Proceedings of the AAAI Conference on Artificial Intelligence, 38(16), 17754-17762. DOI: 10.1609/aaai.v38i16.29728

[14] Lewis, P., et al. (2021). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Transactions of the Association for Computational Linguistics, 9, 199-214. DOI: 10.1162/tacl_a_00363

[15] Borgeaud, S., et al. (2022). Improving Language Models by Retrieving from a Large-Scale Web Dataset.. DOI: 10.48550/arXiv.2207.02199

[16] Gao, L., & Luan, J. (2023). Text-to-Knowledge Generation via Enhanced Pretrained Language Models. Journal of Artificial Intelligence Research. DOI: 10.5555/jair.v79

[17] Liu, Y., Ott, M., Goyal, N., et al. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. ArXiv. DOI: 10.48550/arXiv.1907.11692

[18] Chen, X., & Chen, X. (2023). Cross-Domain Textual Information Extraction in Augmented NLP Models.Proceedings of the International Joint Conference on Artificial Intelligence. DOI: 10.24963/ijcai.2023/475

[19] Izacard, G., et al. (2021). Distilling Knowledge from Generative Question Answering for Multimodal NLP. Computational Linguistics. DOI: 10.1162/coli_a_00451

[20] Zhang, X., & Wu, L. (2023). Transforming Language Understanding with Multilingual Text Extraction and Generation. Artificial Intelligence Journal. DOI: 10.5555/1376879

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

ABOUT IJEAST

International Journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high-quality research papers in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

FOCUS AREAS

- Engineering
- Applied Science
- Technology
- Innovation & Development
- Interdisciplinary Studies

PEER REVIEWED
All submissions are rigorously peer reviewed to ensure quality.

OPEN ACCESS
Free and unrestricted access to research for all.

GLOBAL REACH
Connecting researchers and professionals worldwide.

TIMELY PUBLICATION
We ensure a swift and efficient publication process.

For more information, visit our website
www.ijeast.com

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

editor@ijeast.com
www.ijeast.com
India

2455-2143