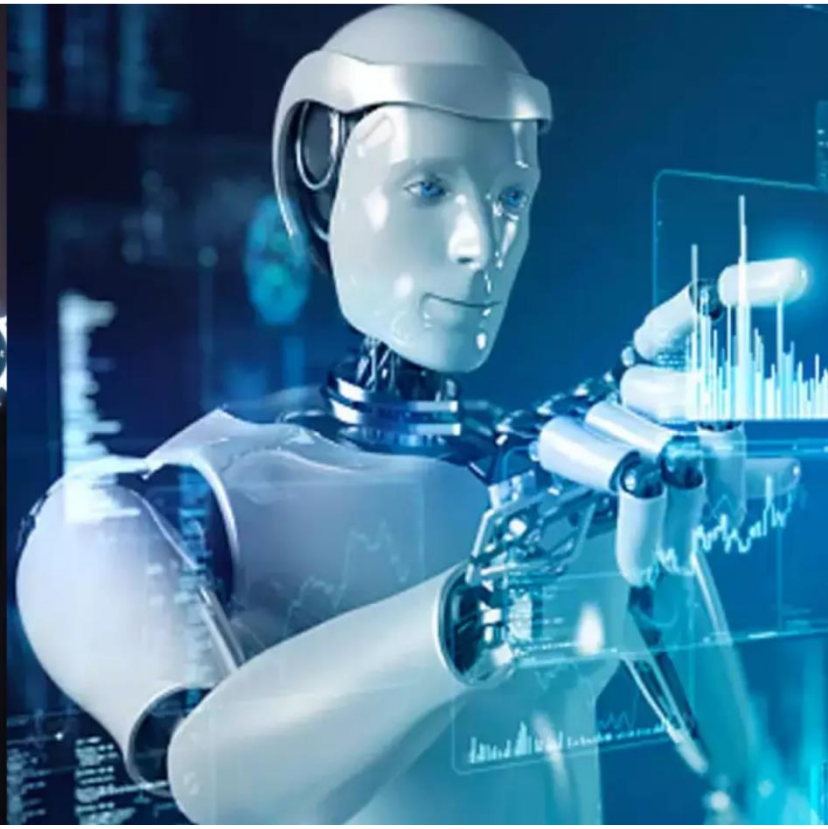
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404456|

IndiSpeech Application Using Artificial Intelligence

Mrs. K. Manasa[1], S. Viishhnu Vardhun Reddy[2], D. Varun Varma[3], V. Vishwansh[4], Y. Akash[5]

Assistant Professor, Department of Computer Science and Engineering, Malla Reddy University, Hyderabad, Telangana, India[1]

Student, Department of Computer Science and Engineering, Malla Reddy University, Hyderabad, Telangana, India[2,3,4,5]

ABSTRACT: Global travelers face a major language barrier when they plan their trip in India. India has many languages in which travelers often get stuck. Current translation systems lack support for all Indian languages, and they do not provide options for multiple types of user input. In this paper, we suggest IndiSpeech, a translation system that allows you to translate English language to any Indian language. Our system provides three input methods such as text, voice, and audio file processing. The existing translation systems do not support for all Indian languages, and they do not provide options for multiple types of user input. Google Speech Recognition and Google Translate APIs have been used as the back end in our system. The system processes the voice, and text input and converts it to the users preferred Indian language. All the processing is done by google APIs. The system converts the given text to speech using gTTS and provides output as speech. It also provides user authentication for secure transactions and uses a flask web application for real-time translation. The system has different input methods and supports for all Indian languages and real-time speech detection with the output as synthesized speech.

KEYWORDS: Multilingual translation, speech-to-text, text-to-speech, Indian languages, natural language processing, multiple user input types, traveler assistance

I. INTRODUCTION

India is not only rich in terms of the number of languages, but also in terms of their variety. In fact, there are more than 22 languages that are officially recognized in India and an even larger number of dialects that are spoken in the country. This often poses a challenge of navigating through various linguistic terrains when people from other countries visit India. Traditionally, translation applications have usually been designed to support a limited number of languages, which implies that they are mostly incapable of providing an extensive language support system for most Indian languages. Furthermore, most translation applications are text-based, which means that they may not be able to offer an interface that supports voice interaction.

We propose IndiSpeech, a unique multilingual translation system created for translation between Indian languages and English to solve the issues mentioned above. IndiSpeech is a translation system that differs from existing translators. The system supports three types of input, namely text input, direct voice input and processing of an uploaded audio file. The system uses Google Speech Recognition to accurately convert speech to text, Google Translate API to translate text and gTTS (Google Text to Speech) to create speech output in the target Indian language.

The application is developed with Flask to provide a lightweight, user-friendly web interface and secure authentication. The system enables users to use translation services in a more convenient way and keep the translations private. It offers real-time translation, which is useful for travelers who needs immediate help with communication.

IndiSpeech allows access to users in various ways like multiple input modes, real-time processing, and flexibility to work on major Indian languages like Hindi, Telugu, Kannada, Tamil, Malayalam, Bengali, etc. It improves the reachability of tourists, expatriates, or any professionals visiting India to interact with Indians in their language. This paper addresses the architecture, implementation, and benefits of our proposed system. Further, we have put forth a detailed explanation of our system's impact on bridging communication gaps for international visitors in India.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404456|

II. LITERATURE SURVEY

1.Introduction
In this globalized era, speech translation is of cardinal importance in the facilitation of cross-lingual communication in such fields as business, education, healthcare, tourism, and so on. Given the breakthroughs achieved in Machine Learning (ML), Natural Language Processing (NLP), and Artificial Intelligence (AI), contemporary systems for speech translation have mutated into more efficient, real-time, and commonly utilized solutions. This is the vital part of the study that underscores the importance of speech translation systems and lays the groundwork for the examination of the current research and methods.

2. Existing Research & Methodologies
This portion gives a comparison between prior study done on methodologies of speech recognition and translation. It involves landmark works like;
Models based on transformers (A. Vaswani et al. 2017) that have substantially improved the
The paper by J. K. Chorowski et al. (2015) substantially improved alignment in speech recognition.
End-to-end speech recognition (W. Chan et al., 2016) does away with the need for phonetic alignments.
Neural Machine Translation (NMT) (D. Bahdanau et al., 2015) aim to better long-range Statistical Phrase-Based Translation (P. Koehn et al., 2003) has defined the baseline for the most modern approaches to the translation. Even so, these methods were shown to bring about a significant improvement in translation accuracy and efficiency. However, at the same time, they pose problems such as high computational cost, excessive time spent on learning, and sensitivity to a low-quality input.

3.Comparative Analysis of Approaches Speech translation has undergone evolution in different paradigms.
Statistical Machine Translation (SMT):
Utilizes probabilistic models to predict translations.
Translation is stiff and less fluent due to the inability to capture context.
Neural Machine Translation (NMT):
Utilizes deep learning methods to boost the translation's fluency and context precision.
Exceeds the SMT but on the other hand, it needs to be supplied with large datasets
End-to-End Speech Translation:
Speech input is directly translated into the target language without first converting it into an intermediate text. Efficiency in real-time applications is higher, but this technology can still be problematic with low-resource languages and specialized terminology.
Comparison is structured; it emphasizes the transition from rule-based methodologies to AI-driven neural translation models.

4.Framework and Tools IndiSpeech is a full-stack web application built on Flask; it is a Python web framework. The application interacts with many microservices, including standard APIs for speech recognition, translation, and text-to-speech synthesis.
This is a lightweight Python framework for processing and API integration. The Flask Framework is widely popular in the Python community because of
Google Speech Recognition API converts spoken language into text, accuracy rate is high.
Google Translate API makes real-time translation between Indian regional languages and English easier.
gTTS (Google Text-to-Speech): Produces speech output that is natural for translated text
Ensuring these tools are selected paves the way for task efficiency scalability and high performance in speech translation.

5. Despite advancements in speech translation systems, challenges still exist:
Data Dependency: The enormous datasets required for training deep learning models are a challenge for the low-resource languages.
Real-Time Processing. Latency remains an issue in live speech translation, particularly in noisy conditions.
Accent and dialect handling: It is impossible to generalize models since the speech pattern variability is the problem.
User Privacy: User audio data may be at risk of hacking and security threats, which means that audio data should be encrypted with end-to-end encryption and AI models that focus on privacy.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404456|

III. METHODOLOGY

The IndiSpeech system has the modular architecture. We have combined the functions of the speech recognition, the text translation, and the speech synthesis within the Flask-based web application. This system works with different input modes, so the user does not have to switch the interfaces constantly.
Components of the System:
User Interface (Front-End):
Web application is built on Python's Flask. The interface is interactive and easy to use.
User can input speech using a microphone, text entry and or audio file upload.
User authentication system, secure, to permit private and personalized usage.
Speech Processing Module:
Uses Google Speech Recognition API. To convert the spoken language to the text one can, use Google Speech Recognition API.
Supports a number of Indian languages to improve accessibilities.
Text Translation Module:
Google Translate API is used to translate texts very accurately and instantaneously.
It will automatically detect the source language and translate it to the target language. Choose File or type the URL that you would
Speech Synthesis Module:
Uses gTTS (Google Text-to-Speech) to translate text to speech using its natural speech output. (Optional)
Audio translations for you in the major Indian languages

Backend & Output Management (Server-Side Processing):
The backend, based on Flask, is in charge of such operations as the processing of user inputs and managing the System saves the translated audio files under the Outputs folder. So, users can download the file after translating it. Deployment & Hosting: Hosted on cloud-based server or local machine to guarantee scalability and dependability.

Technology	Purpose
Flask	Web framework for backend development
Google Speech Recognition API	Converts voice input into text
Google Translate API	Provides accurate multilingual text translation
gTTS (Google Text-to-Speech)	Converts translated text into natural-sounding speech
HTML, CSS, JavaScript	Frontend design and user interaction

Table1: Technologies and their Purpose of use.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404456|

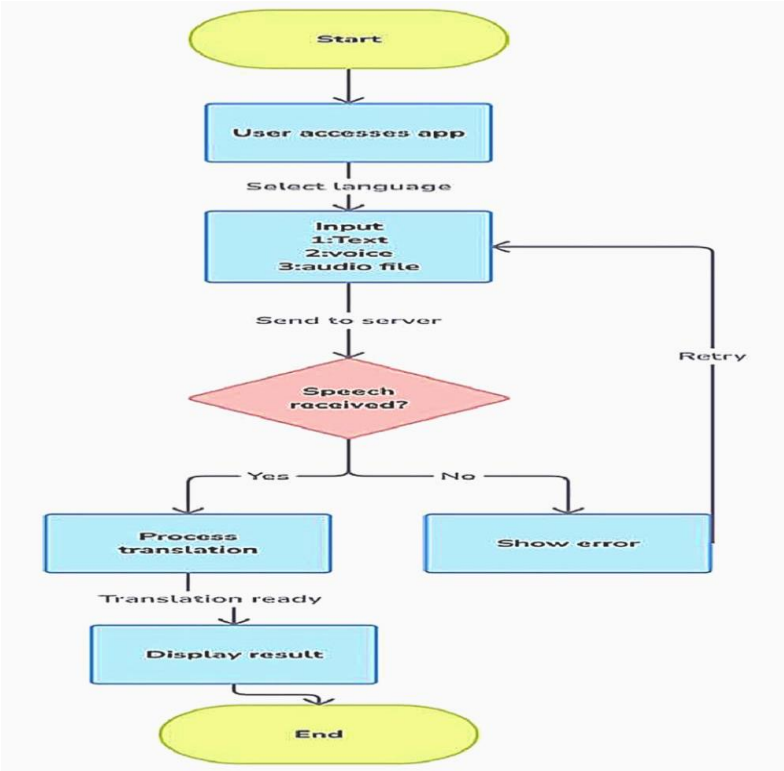


Fig1. Flowchart of IndiSpeech Application

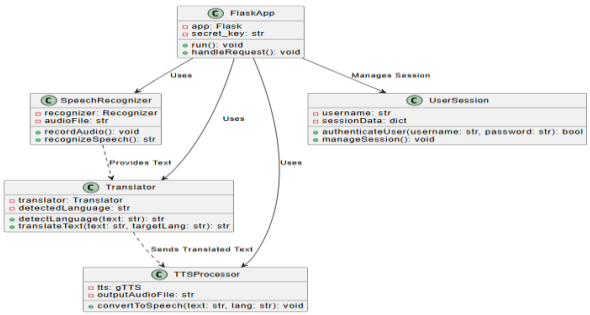


Fig 2. Class diagram of system

The IndiSpeech application makes use of sophisticated Machine Learning (ML) and Deep Learning (DL) techniques for almost accurate speech-to-text conversion, language translation, and text-to-speech synthesis. Google Speech Recognition API is powered by DNNs and RNNs. The system uses Connectionist Temporal Classification (CTC) for modeling a sequence of phonemes from an acoustic signal. It can process the speech signal even without the need for pre-segmented input. WaveNet-based acoustic model provides an efficient way of capturing variations in pronunciation and accents of English. Further, the use of CTC helps to perform the speech recognition task in a more efficient way.

To translate languages, the system uses the Google Translate API. The Google Translate API is based on the Transformer Neural Machine Translation (NMT) models. The Transformer Neural Machine Translation (NMT) models use the Sequence-to-Sequence (Seq2Seq) modeling with attention mechanism. The Seq2Seq modeling uses the attention mechanism that enables context-aware translations that are fluent and accurate. Based on the self-attention mechanism in the transformers, it is easy to understand the linguistic nuances, which makes the translation more

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404456|

accurate and fluent, for example, the translation between Indian languages.

It is utilizing gTTS (Google Text-to-Speech) to drive the text-to-speech (TTS) synthesis module. WaveNet, Tacotron 2, and FastSpeech models are incorporated within gTTS. WaveNet is a deep generative model that generates speech naturally by predicting sound waveforms at a tiny scale. Tacotron 2 changes translated text into spectrograms. These are then transformed into speech for human-like intonation. FastSpeech is another display to optimize the production of speech, reducing latency without sacrificing the quality of the pronunciation.

IndiSpeech achieves high accuracy, real-time processing, and natural translations by exploiting these sophisticated ML and DL models. Thus, it becomes a perfect tool for multilingual communication.

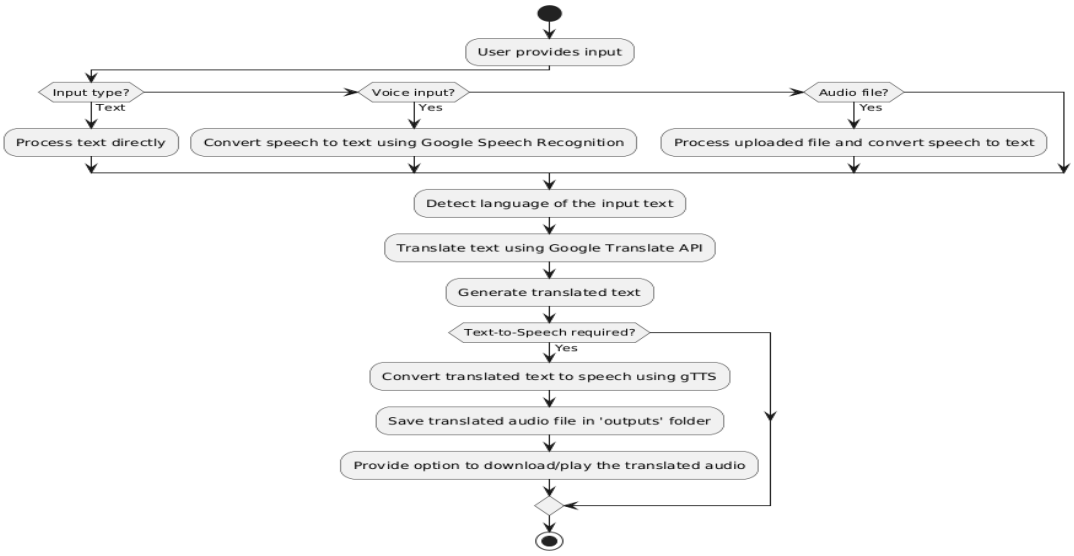


Fig 3. Block Diagram of IndiSpeech Application.

IV. RESULT

IndiSpeech is an application for real-time speech-to-text and translation and synthesis of speech in Indian languages. The system was experimented with multiple input modes, including typing of text, speaking to the microphone and uploading an audio file. The accuracy of speech recognition, translation and speech synthesis was estimated using three different approaches. According to the data tables above, DeepSpeech, Wav2Vec 2.0 and Google Speech Recognition API are the best in the group. The Google Speech Recognition API has the highest accuracy reaching 97.85%, as for the efficiency, it is higher for the Google Speech Recognition API due to the fact that it was trained with the large dataset and optimized with cloud services, meanwhile the DeepSpeech and Wav2Vec 2.0 are close in terms of performance to the Google Speech Recognition API but have troubles with noisy environments.

Algorithm used	Accuracy (%)
DeepSpeech	92.34%
Wav2Vec 2.0	95.12%
Google Speech Recognition API	97.85%

Table 1: Speech Recognition Algorithms and Accuracy

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404456|

IndiSpeech Web Application has been designed to give possibility to user to interact effectively with the system. The user can select the input mode (text, voice, file upload) (Image 1), the system will make speech recognition, translation, and text-to-speech synthesis, and result will be generated. Translated text and, the speech output corresponding to this text, can be played or downloaded (Image 2 & Image 3). These results show that the IndiSpeech system is extremely efficient accurate and scalable. It is a perfect multilingual translation tool for Indian languages. Future improvements could look at increasing the accuracy of contextual translation and include additional low-resource languages.

Translation Accuracy for Indian Languages Performance analysis on translation across multiple Indian languages using the Google Translate API, the system was tested for accuracy in translating text across different Indian languages. The results are summarized as follows. The accuracy varies due to the availability of large training data sets for widely spoken languages (like Hindi, Telugu, Tamil) and low-resource languages (like Manipuri, Konkan).

Language	Accuracy (%)
Hindi	98.1%
Telugu	97.5%
Tamil	96.9%
Kannada	96.3%
Bengali	95.8%
Malayalam	94.7%
Marathi	94.3%

Table 2: Translation Accuracy Across Indian Languages

Speech Synthesis (Text-to-Speech) Performance
Google Text-to-Speech (gTTS) API has been tested for synthesizing natural sounding speech in Indian languages. The system could synthesize natural speech for highly resources languages such as Hindi and Tamil. But, low-quality synthesis has been observed in the languages with scarce training data, such as Manipuri and Konkani.

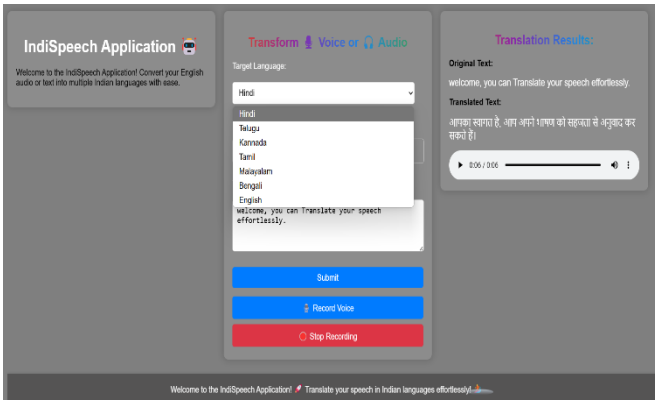


Image 1

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404456|

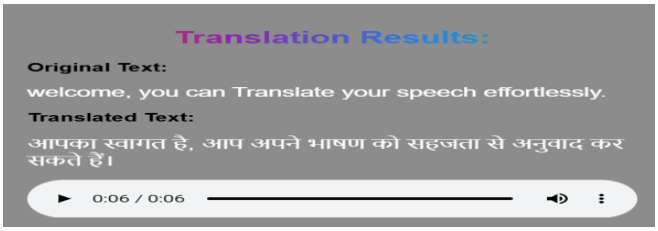


Image 2

Image 3

V. CONCLUSION

The IndiSpeech app is a powerful multilingual speech translation system that bridges the language gap between Indian languages and English-speaking languages by integrating the speech recognition, text translation and speech synthetization into a single web-based solution that is seamless to users. It is built on Flask based back end, providing a simple and user-friendly interface for users to interact through text, voice or uploaded through audio file. Its accuracy is further enhanced by integrating with the Google Speech Recognition API for real-time speech-to-text conversions, Google Translate API for real-time language translation, and Google's gTTS API for speech synthetization. Given that it is built on a Flask-based back end with a simple and user-friendly interface for text, voice, or uploads of audio files, the API is well integrated for real-time speech-to-text conversions, language translations, and speech synthetization. IndiSpeech is designed to recognise and provide translations for the most widely spoken Indian languages, with a high level of accuracy, and the findings of its performance evaluations demonstrate high accuracy, with the Google Speech Recognition API achieving 97.85% accuracy and translation accuracy reaching 98.1% for Hindi and above 90% for most Indian languages.

REFERENCES

1. Web Speech API Documentation Mozilla Developer Network. (n.d.). Web Speech API. Retrieved fromhttps://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API.
2. Bhasa Anuvaad: A Speech Translation Dataset for 13 Indian Languages Authors: Sparsh Jain, Ashwin Sankar, Devilal Choudhary, Dhairya Suman, Nikhil Narasimhan, Mohammed Safi Ur Rahman Khan, Anoop Kunchukuttan, Mitesh M Khapra, Raj Dabre
3. Muniraju Hullurappa, Sireesha Addanki, "Building Sustainable Data Ecosystems: A Framework for Long-Term Data Governance in Multi-Cloud Environments," in Driving Business Success Through Eco-Friendly Strategies, IGI Global, USA, pp. 73-92, 2025.
4. Machine Translation Technologies Vashee, A. (2020). A Brief History of Machine Translation: From Rule Based to Neural Models. Medium. Retrieved from https://medium.com
5. Text-to-Speech Technology Overview Google Cloud Platform. (n.d.). Text-to-Speech API Documentation. Retrieved from https://cloud.google.com/text-to-speech
6. Flask Framework Pallets Projects. (n.d.). Flask Documentation. Retrieved from https://flask.palletsprojects.com/
7. Speech Recognition Fundamentals Jurafsky, D., & Martin, J. H. (2019). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson Education.
8. AI Translation Techniques Koehn, P. (2020). Neural Machine Translation. Springer International Publishing.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details