Tribhuvan University

Orchid International College

A Final Year Internship Report

On

“Artificial Intelligence/Machine Learning”

At

Wiseyak Solution Pvt. Ltd.

Under the Supervision of

Pawan Niroula

Lecturer

Nitisha Timalsina (20879/076)

Submitted To:

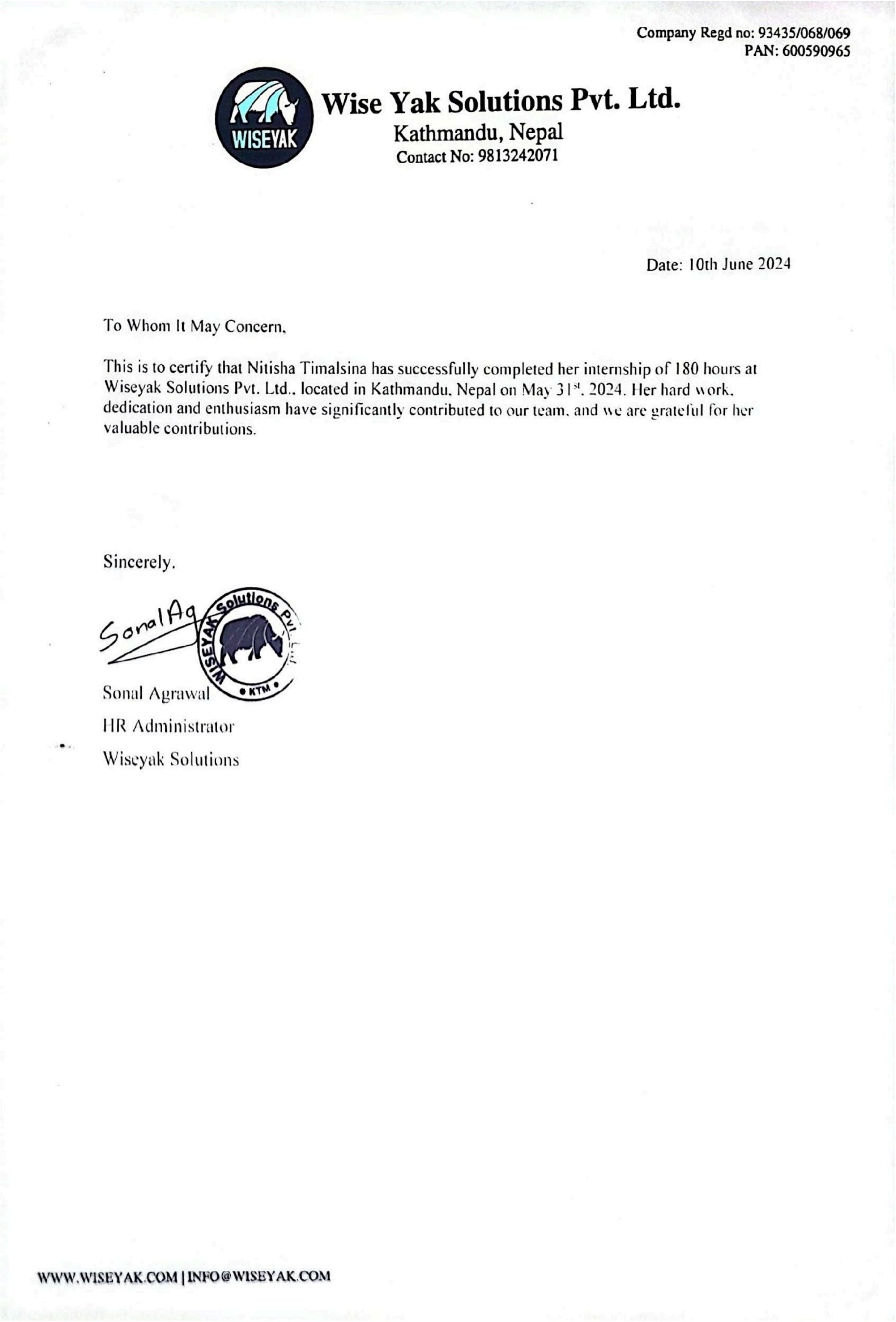
Department of Computer Science and Information Technology

Orchid International College

In partial fulfillment of the requirement for the Bachelor Degree in Computer Science and Information Technology

June, 2024

COMPLETION LETTER



Company Regd no: 93435/068/069
PAN: 600590965

WISEYAK

Wise Yak Solutions Pvt. Ltd.
Kathmandu, Nepal
Contact No: 9813242071

Date: 10th June 2024

To Whom It May Concern,

This is to certify that Nitisha Timalsina has successfully completed her internship of 180 hours at Wiseyak Solutions Pvt. Ltd., located in Kathmandu, Nepal on May 31st, 2024. Her hard work, dedication and enthusiasm have significantly contributed to our team, and we are grateful for her valuable contributions.

Sincerely,

SonalAg

Wiseyak Solutions Pvt. Ltd.
KTM

Sonal Agrawal
HR Administrator
Wiseyak Solutions

WWW.WISEYAK.COM | INFO@WISEYAK.COM

ORCHID INTERNATIONAL College
[TRIBHUVAN UNIVERSITY AFFILIATE]

SUPERVISOR’S RECOMMENDATION

I hereby recommend that the report prepared under my supervision by Nitisha Timalsina (TU Exam Roll No. 23879/076 entitled “**Artificial Intelligence/Machine Learning**” at **Wiseyak Solution Pvt. Ltd.** in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Information Technology be processed for evaluation.

……………………..…….

Pawan Niroula
Lecturer, Department of CSIT
Orchid International College
Bijayachowk, Gaushala

ORCHID INTERNATIONAL College
[TRIBHUVAN UNIVERSITY AFFILIATE]

CERTIFICATE OF APPROVAL

This is to certify that this project prepared by Nitisha Timalsina (TU Exam Roll Number: 23879/076) entitled “**WISEYAK SOLUTION PVT. LTD**” in partial fulfilment of the requirements for the degree of B. Sc. in Computer Science and Information Technology has been well studied. In our opinion it is satisfactory in the scope and quality as a project for the required degree.

Pawan Niroula
Supervisor,
Full Time Faculty
Orchid International College
Bijayachowk, Gaushala

Er. Dhiraj Kumar Jha
Head Of Department,
Orchid International College
Bijayachowk, Gaushala

External Examiner
Central Department of Computer Science and IT,
Tribhuvan University,
Kirtipur, Nepal

ACKNOWLEDGEMENT

I would like to honor and express our deepest appreciation to everyone who has assisted us and express my sincere gratitude to the supervisor of this internship program **Er. Dhiraj Kumar Jha,** Full Time Faculty, Department of Computer Science for his constant support, guidance and valuable suggestions throughout the internship program. Also, I would extend my sincerest gratitude and appreciation to **Suraj Prasai**, mentor of my internship program.

I would like to thank **Wiseyak Solutions Pvt. Ltd.** for providing me a wonderful opportunity at the organization. I am grateful for my seniors **Er. Kishan Kumar Sharma** and **Er. Neelam Karki** from the wiseyak team for continuous support and providing me encouragement during my internship period.

I am also grateful to the Department of Computer Science and Information Technology, Orchid International College, for providing a wonderful internship opportunity and commitment from teachers for providing a proper environment for facing challenges during the project.

Sincerely,

Nitisha Timalsina (23879/076)

Abstract

This report presents an account of the author's internship experience at Wiseyak Solution Pvt Ltd, where the author engaged in several projects centered around Artificial Intelligence (AI) and Machine Learning (ML). The internship spanned 180 hours, during which the author contributed to the development and enhancement of AI/ML models and solutions tailored to address various industry-specific challenges. The responsibilities included data preprocessing, model training and evaluation, and the deployment of machine learning algorithms. A significant portion of the work involved collaborating with cross-functional teams to integrate AI capabilities into existing systems, optimizing algorithms for improved accuracy and efficiency, and participating in research and development activities to innovate new AI-driven features. Through these projects, the author gained hands-on experience with cutting-edge technologies and tools such as PyTorch, scikit-learn and various learnings about the large language models, while also developing a deeper understanding of AI/ML principles and best practices. The report elaborated on specific projects undertaken, including the implementation of natural language processing (NLP) techniques for text analysis, the creation of predictive models for data-driven decision-making. Each project was discussed in terms of objectives, methodologies, outcomes, and the technical challenges encountered and addressed. Through the internship project, the author significantly enhanced technical expertise in AI/ML, provided valuable industry insights, and fostered a collaborative and innovative approach to problem-solving in the field of artificial intelligence.

Keywords: Artificial Intelligence, Machine Learning, Natural Language Processing, Large Language Models,

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
API	Application Program Interface
BERT	Bidirectional Encoder Representation
CNN	Convolutional Neural Network
CSV	Comma Seperated Values
CV	Computer Vision
EDA	Exploratory Data Analysis
GPT	Generative Pre-trained Transformer
IDE	Integrated Development Environment
IT	Information Technology
JSON	Javascript Object Notation
LLaMA	Large Language Model Meta AI
LLM	Large Language Model
ML	Machine Learning
NLP	Natural Language Processing
R&D	Research and Development
RNN	Recurrent Neural Network
VAD	Voice Activity Detector

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 - INTRODUCTION

1.1 Introduction

Internship is a period of practical work experience for the individuals (students or recent graduates) who are undertaken into an organization or the company. During the internship period, interns get exposure to the tasks, responsibilities and day-to-day workings, develop networking to professional in an organization. As a BSc.CSIT course, student must fulfill their internship with minimum period of 10 weeks or 180-hour period for an internship.

Internship provides practical works to the individual such that they can align their skills with the working domain.

Artificial Intelligence can be generally categorized along two main dimensions: one based on thought processes and reasoning or behavior, and the other on fidelity to human performance or ideal (rational) performance. Rationality refers to the quality of being logical or reasonable. Consequently, four distinct approaches to AI have been recognized and accepted. The human-centered approach involves examining and hypothesizing about human behavior, while the rational approach combines principles from mathematics and engineering. Artificial Intelligence includes four basic approaches Thinking Humanly: The cognitive modeling approach. Acting Humanly: The Turing Test Approach, Thinking Rationally: The "laws of thought" approach, "Acting Rationally": The rational agent approach.

Machine Learning is a part of Artificial Intelligence, which uses various statistical algorithms to automatically learn from data and improve from experience without being programmed manually. The machine finds the correlation between inputs or between inputs and outputs and recognizes pattern within the data which it uses to predict or cluster unseen data in future. The main thing is that the machine learns all the rules on its own without any human intervention. primary tool for understanding the features. In image captioning, images are taken as input and their respective descriptions are provided as output. Sentiment analysis is the process of getting the emotional tone or sentiments expressed in a piece of text. The input of sentiment analysis is text description and output are sentiment (Positive, Negative or Neutral).

The forthcoming developments in AI and machine learning are poised to revolutionize numerous sectors, fostering efficiency gains and unlocking fresh possibilities, spanning healthcare to finance. Interning in this vibrant field affords the opportunity to immerse oneself in leading-edge technologies, participating in inventive initiatives that shape the trajectory of intelligent systems. This role presents a distinct avenue for skill acquisition and impactful contributions within a dynamic and swiftly evolving landscape.

1.2 Problem Statement

In the rapidly evolving landscape of AI and ML, the pressing need for more ML engineers is met with several obstacles. Primary among these is the scarcity of individuals equipped with the necessary ML skills, including a deep grasp of algorithms, techniques, and tools. Furthermore, the intricate nature of ML systems demands a strong foundation in mathematics, statistics, and computer science, alongside proficiency in fields like data science and software engineering. Staying updated with the latest advancements in AI/ML is an ongoing challenge, as is navigating the ethical considerations and biases inherent in ML algorithms. Lastly, the rapid pace of technological progress means that the skills acquired by ML engineers today may quickly become obsolete, underscoring the importance of continuous learning and professional growth. Tackling these challenges is crucial for meeting the escalating demand for ML engineers and effectively harnessing AI's potential to solve complex real-world problems.

1.3 Objectives

The internship aims to provide individuals with practical exposure in their chosen field, enhancing their professional development through the application of academic learning. Interns will have the opportunity to acquire various skills and technologies, including professional software development, real-world problem-solving, teamwork, documentation, communication, and understanding organizational culture. Consequently, the primary goal of the internship can be categorized into two parts: Internship Objectives and Project Objectives.

1.3.1 Internship Objectives

The main objective of the internship is to gain experience and professionalism for seeking a role as ML engineer, and practical exposure to real-world challenges in the dynamic field of AI and ML. With this internship one will gain hands-on experience in addressing prevalent industry hurdles, such as the shortage of skilled ML professionals, the intricacies of ML systems, ethical dilemmas, and the necessity for continuous learning. Through active engagement with these challenges, internship will cultivate a comprehensive understanding of the field. It encourages students to contribute towards innovative solutions, fostering growth and innovation in AI/ML. Ultimately, the internship seeks to equip with the skills and mindset needed to positively impact the responsible advancement of AI and ML technologies across various domains. Here are some main objectives of the internship:

- To incorporate practical knowledge acquired from the classroom.
- To get hands on practical experience with the respective subject.
- To understand and learn real-time technical, organizational culture, managerial and life experience required at job.
- To get insights into career opportunities through observation, interaction and work experience through the organization.

1.3.2 Internship Project Objectives

Organization provides an opportunity to work in project which helps intern to understand working culture and provide the familiarity with real world projects. Here are some of the objectives of the internship project:

- To get familiarity with Artificial Intelligence and Machine Learning tools and techniques.
- To get knowledge related to ML frameworks for natural language processing, ML algorithms and research on different ML models.
- To get acquainted with different tools and frameworks like: PyTorch, Scikitlearn, LLM, V-A-D.

1.4 Scope and Limitation

The project is built to the purpose of incorporating voice chats to the existing chatbot. The user will have to ask information about the particular field and responses and generated through the chatbot. This includes both text-to-speech and speech-to-text models. Through the overall system, user will get detailed information related to the client specific questions.

1.5 Report Organization

The entire document had been divided into four parts, which are listed below:

Chapter 1: Introduction

This section includes the introduction to internship and basic overview of the problem statement, objectives to the project and scope, limitation of the project.

Chapter 2: Organization details and literature review

This section contains details related to the organization like: basic introduction of organization, organizational domains, working domains, description of intern work and literature reviews related to the internship domains like: research papers, articles and resources explored during the implementation of the project.

Chapter 3: Internship activities

It consists of roles and responsibilities of the intern during the internship period, weekly logs, detailed description of the project. All the technical activities performed in the project during the internship period are included in this chapter.

Chapter 4: Conclusion and learning Outcomes

This chapter is the final chapter of the document which includes the conclusions drawn during the work in the project.

CHAPTER 2 - ORGANIZATION DETAILS AND LITREATURE REVIEW

2.1 Introduction to Organization

Wiseyak is a healthcare technology company that harnesses the power of AI and ML to enhance the healthcare industry. Its healthcare data platform, WiseMD, is tailored to capture standardized and interoperable medical data, offering invaluable insights for clinical decision support and medical diagnosis. Additionally, Wiseyak provides AI-related consulting services to many other companies in Nepal. It is a US based healthcare startup established in 2019. The organization is currently located at Naxal, Kathmandu. The team includes learned professionals and engineers for delivering standard products.

Wiseyak team consists of software developers, machine learning engineers and experts with years of experience. Wiseyak aspires to develop decision support systems, customized to each client's based on their requirements.

Table 2.1 Contact Details of Organization

Location	Naxal Bhatbateni, Kathmandu, Nepal
Website	www.wiseyak.com
Email	info@wiseyak.com
Phone	01-4412472

2.2 Organizational Hierarchy

An organization is composed of group of individuals who are working together to achieve common mission, vision, goals and objective. These four key components could range from business objectives, societal missions, research advanced to any other specific purpose that brings people together.

Currently, the team size of Wiseyak comprises with 31 members divided into two departments; Software development department and the AI/ML Department where software development department is handled by the project manager and the AI/ML team works under the AI/ML team lead. Here is the general organizational hierarchy of the Wiseyak solutions Pvt. Ltd.:

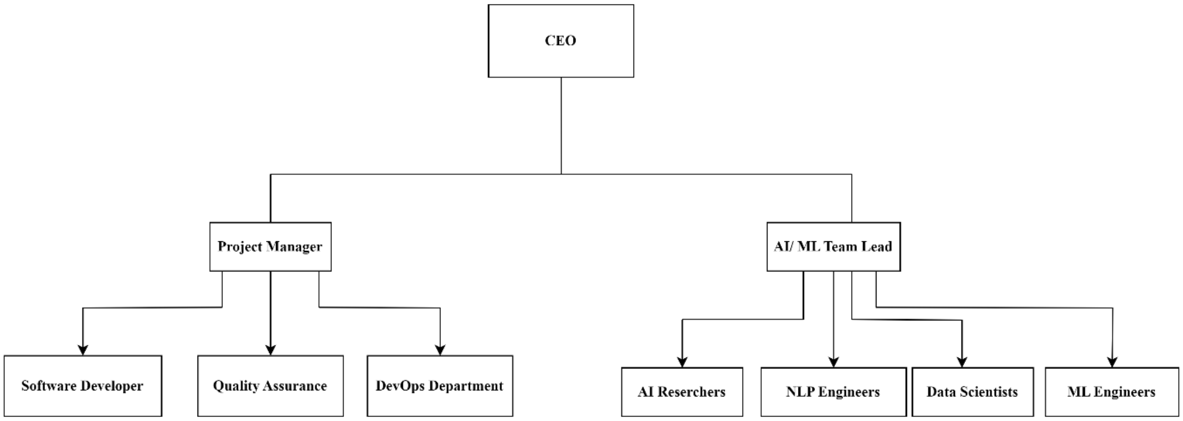


Figure 2.1 Organizational Hierarchy of Wiseyak Solutions

2.3 Working Domains of the Organization

Wiseyak Solution Pvt Ltd specializes in the development of a diverse array of software applications. Their expertise encompasses the creation of sophisticated web applications, the development of mobile applications, and the implementation of AI-driven electronic medical record (EMR) systems. They also focus on building communication-driven decision support systems designed to enhance organizational efficiency and decision-making processes. Moreover, Wiseyak is actively involved in AI research, contributing to advancements in artificial intelligence technologies and their practical applications across various industries. Here are some of the working domains of the organization:

2.3.1 Web Application Development

Wiseyak has extensive experience on EMR based web applications. The organization properly plan and arrange, design, develop, deploy the interactive EMR applications. They are also focused on large web applications with dedicated employees.

2.3.2 Mobile Application Development

Wiseyak offers mobile application development. The organization provides cross platform applications supporting both in iOS and Android.

2.3.3 AI Research

Wiseyak Solutions has experience on providing research in AI and data related fields. Different machine learning models and data engineering technologies that are integrated with web applications and mobile applications are researched on. The organization primarily focuses on medical researches, and NLP as their main domains of AI.

2.3.4 Consulting

Wiseyak Solution provides expert assistance suited to unique needs and delivers comprehensive consulting services to a range of organizations and corporations. In order to ensure the smooth integration of cutting-edge AI systems, their consulting services concentrate on integrating AI-driven solutions into the organization's development strategies.

2.4 Description of Intern Department

2.4.1 Selection and Placement

Wiseyak solution has a proper procedure to hire and manage employees. The selection of interns is a two-way process. First, the student must match the company’s job opening with their area of interest, and second, the company must be open to hiring the undergraduate as a student or employee. The applicants get in the organization through multiple rounds of interview after the screening of CV. After the CV screening and successful interview processes, the student is hired for the desired position. Interns in the organization are involved in various activities performed by the company, showcasing their knowledge and skills by getting involved in different projects and aspects of the organization.

For carrying out evaluation of interns in internship period, project manager and team leads are responsible. The mentorship is provided by the ML team lead after the interns are onboarded. A daily or weekly task must be. At the end of three-month internship period, mentors evaluate the performance of an intern. Later on, if the overall performance of intern is appealing then, they are hired in the junior position based on their department.

2.4.2 Duration

The duration and relevant details of the internship are as follows:

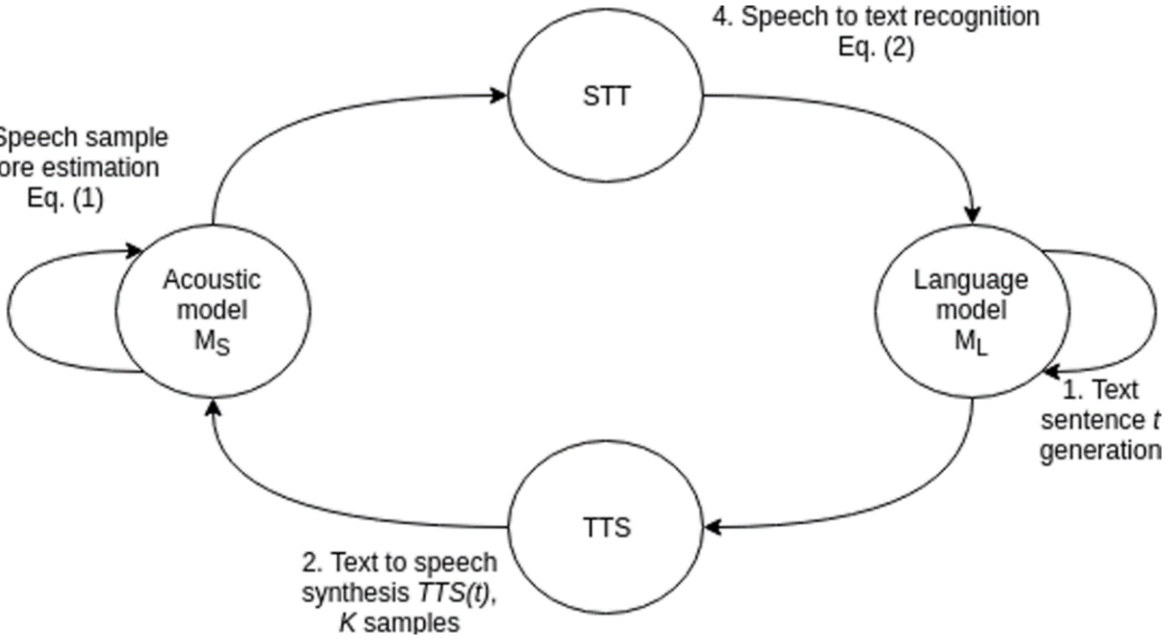
Table 2.2 Internship Duration

Internship Position	AI/ML Intern
Start Date	Feb 29, 2023
End Date	May 29, 2023
Internship Period	3 Months
Hours per day	6.5 Hours (11 AM – 5:30 PM)
Mentor	Suraj Prasai, AI/ML Team Lead

2.5 Literature Review

Text-to-speech (TTS) and speech-to-text (STT) technologies have seen significant advancements in recent years, making them integral to the development of chatbots that facilitate natural and effective human-computer interactions. These technologies are pivotal for creating chatbots that can both comprehend [1] spoken language and respond in a human-like manner. This review examines recent progress, applications, and ongoing challenges in TTS and STT for chatbots.

The evolution of TTS technologies has been marked by a shift from early methods to more sophisticated approaches. Initially, TTS systems relied on concatenative synthesis, which involved piecing together pre-recorded snippets of speech. This method often resulted in unnatural and robotic-sounding speech. Modern TTS systems, however, utilize neural network-based approaches that generate more natural and fluid speech. For instance, Google's WaveNet, developed by DeepMind, uses a deep generative model for raw audio waveforms, producing speech that closely mimics human patterns. Another significant development is Tacotron, and its successor Tacotron 2, [2] which convert text to mel-spectrograms and then to audio waveforms, resulting in highly intelligible and natural-sounding speech. The applications of TTS technologies are diverse and impactful. In assistive technologies, TTS is crucial for devices designed to aid individuals with visual or speech impairments. Virtual assistants like Siri, Alexa, and Google Assistant rely on TTS to provide information and perform tasks based on user commands. Additionally, businesses deploy TTS in automated customer service systems to offer 24/7 support, enhancing user experience and operational efficiency.



Dual supervised learning for non-native speech recognition

Figure 2.2 TTS and STT Process

Speech-to-text technologies have also evolved significantly, moving from early methods to cutting-edge techniques. Early STT systems were based on Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs), which required extensive labeled data for training and often yielded limited accuracy. The adoption of deep learning, particularly recurrent neural networks (RNNs) and transformer models, has revolutionized STT systems, enhancing both their accuracy and efficiency. For example, Mozilla's DeepSpeech uses RNNs and a character-level model to achieve high accuracy in speech recognition tasks. Similarly, transformer-based models like Facebook AI Research's Wav2Vec 2.0 leverage self-supervised learning to better understand speech patterns with less labeled data.

Voice Activity Detection (VAD) algorithms are essential components in various audio processing applications, including telecommunication systems, speech recognition, and assistive technologies. The primary goal of VAD is to distinguish between speech and non-speech segments in an audio stream, which is critical for enhancing the performance of subsequent speech processing tasks. Early VAD methods relied on energy thresholding techniques, which were simple but often failed in noisy environments due to their sensitivity to background noise and variations in speech intensity.

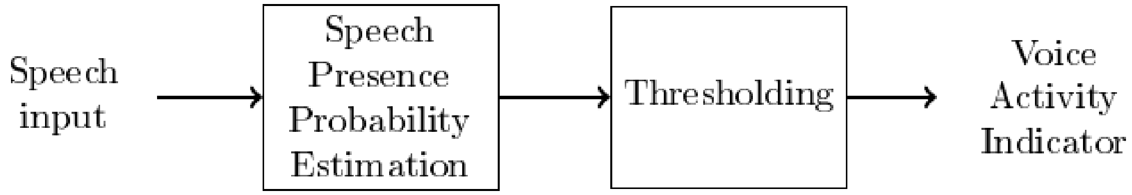


Figure 2.3 Voice Activity Detector

More sophisticated approaches introduced statistical models such as Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs), which improved robustness by incorporating probabilistic frameworks to model the speech and non-speech characteristics. These methods, however, required extensive training data and computational resources. The advent of machine learning, particularly deep learning, has significantly advanced VAD algorithms. Techniques employing deep neural networks (DNNs) and recurrent neural networks (RNNs) have demonstrated remarkable improvements in accuracy and noise robustness. For instance, models like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks are capable of capturing temporal dependencies and complex patterns in audio signals, thereby enhancing the detection of speech in diverse acoustic environments.

In recent developments, self-supervised and unsupervised learning methods have been explored to further reduce the dependency on labeled data. These approaches leverage large volumes of unlabeled audio data to learn useful representations, which can be fine-tuned with minimal supervision for VAD tasks. Additionally, integrating VAD with other speech processing technologies, such as speech-to-text (STT) and text-to-speech (TTS) systems, has become a focal point. For instance, in voice-activated assistants, VAD ensures that only relevant speech is processed by STT systems, thereby improving efficiency and accuracy.

Despite these advancements, challenges remain in achieving reliable VAD performance across various conditions, such as different noise levels, speaker variations, and acoustic environments. Ongoing research is focused on developing more adaptive and resilient algorithms, utilizing hybrid approaches that combine traditional signal processing techniques with advanced machine learning models. The continuous improvement in computational power and the availability of large datasets are expected to drive further innovations in VAD technology, making it more robust and versatile for real-world applications.

CHAPTER 3 - INTERNSHIP ACTIVITIES

3.1 Roles and Responsibilities

During the three-month internship at Wiseyak Solutions Pvt. Ltd., the role and responsibilities of the AI/ML intern encompass a diverse range of tasks essential for professional growth and contribution to real-world projects in artificial intelligence and machine learning. The primary responsibility lies in learning and development, where active participation in training sessions and workshops is crucial for enhancing understanding of AI/ML concepts, algorithms, and techniques. Additionally, tasks involve data collection and preprocessing, including gathering relevant datasets and cleaning and transforming data for analysis. Developing a reusable code for future purpose.

- Actively participate in training sessions and workshops to enhance your understanding of AI/ML concepts, algorithms, and techniques. Stay updated with the latest advancements in the field by conducting research and self-study.
- Assist in the collection, cleaning, and preprocessing of data for machine learning models. This involves gathering relevant datasets, performing data cleaning and transformation, and preparing the data for analysis. Writing scripts for google place transformation.
- Collaborate with the team to develop machine learning models for various applications. This includes selecting appropriate algorithms, building and training models, and fine-tuning them for optimal performance.

- Conduct thorough evaluation and testing of the developed models to assess their accuracy, robustness, and efficiency. Analyze the results and provide insights for model improvement.

3.2 Weekly Log

Weekly logs are kept in this area during the 12-week internship duration. The report's appendix part contains the signed log reports. Here is the weekly log providing the details of the responsibilities given by the mentor, activities performed by the author, the intermediate progress or observation found during the internship period:

Table 3.1 Weekly Log

Time Period	Activities Performed
First week	**Responsibilities:** ● Learning the book Practical Statistics for Data Science. Pdf from, https://github.com/gedeck/practical-statistics-for-data scientists?tab=readme-ov-file ● Chapter-1 Exploratory Data Analysis. Chapter-2 Data Sampling and Distribution.To explore different python packages. **Activities Performed:** ● Made notebooks and successfully discovered new learning materials. ● Research on different NLP tools like: NLTK, IBM Watson. **Observations:** ● Use of NLP in the various text and speech bots.
Second Week	**Responsibilities:** 1. Learning the Practical Deep Learning using fastai course https://course.fast.ai/.

	● To explore foundation of NLP. **Activities Performed:** ● Research and practice coding using PyTorch. ● Data annotation and sentiment analysis for NLP. **Observations:** ● Usage of PyTorch library for NLP .
Third Week	**Responsibilities:** ● Introduction of Language Modeling and NLP. **Activities Performed:** ● Learned about Language translation models. ● Learned to code in the PyTorch framework. **Observations:** ● Incorporating LLMs using PyTorch framework.
Fourth Week	**Responsibilities:** ● Research on Convolutional Neural Network and Recurrent Neural Network to incorporate in the ongoing medical research. **Activities Performed:** ● Learned different research papers and performed researches. **Observations:** ● Acquiring knowledge reading various research papers.
Fifth Week	**Responsibilities:** ● To explore Voice Activity Detector.

	● Learning to use the V-A-D Algorithm. ● Learning to integrate the V-A-D Algorithm in the ongoing project. **Activities Performed:** ● Learned about integration of the V-A-D library in the project. **Observations:** ● Discovered about the version control and collaboration processes in the company. ● Deployment process in the production.
Sixth Week	**Responsibilities:** ● Using Voice-Activity-Detector in the organization's ChatBot. **Activities Performed:** ● Tuning of the VAD in the ChatBot. ● Waiting time of the speech in the ChatBot using VAD. ● Integrated VAD in the ChatBot using React Framework. **Observations:** ● Threshold must be tuned and tested. ● Same mechanism to be used in the next project.
Seventh Week	**Responsibilities:** ● Waiting time of the speech in the ChatBot using VAD. ● Threshold tuning of the speech in the ChatBot. **Activities Performed:** ● Overcome with the problems of thresholds in the ChatBot.

	● Integrated VAD in the ChatBot using React Framework in two projects. **Observations:** ● Error solving and understanding. ● Successful testing
Eighth Week	**Responsibilities:** ● To improve the mic recording button. ● Using V-A-D audio sequence management. **Activities Performed:** ● Resolving the pre speech issues. **Observations:** ● Problems with pre-speeches in the speech bot. ● Tuning the mic and speech in the chatbot.
Ninth Week	**Responsibilities:** ● To implement automatic speech recognition (ASR) and Nepali speech translation in the upcoming projects. **Activities Performed:** ● Development and understanding of the ASR models like Whisper. **Observations:** ● Finding different models for ASR like Whisper in hugginface. ● Leaning about different NLP models.

	● Using Ollma for chatbot text-to-text models. ● Overall completion of the project.

3.3 Project Description

3.3.1 Introduction to Project

The Voice Activity Detection (VAD) and Chatbot projects represent critical advancements in the field of conversational artificial intelligence. The VAD project focuses on developing algorithms that can accurately distinguish between speech and non-speech segments in audio signals. This technology is essential for improving the efficiency and accuracy of voice-activated systems by ensuring that the system responds only when actual speech is detected, thereby reducing false triggers and enhancing user experience.

The Chatbot project aims to create an intelligent conversational agent capable of understanding and responding to user inputs in both text and speech formats. Integrating the VAD functionality, the chatbot can handle voice inputs more effectively, making it versatile and user-friendly. The chatbot is designed to support multilingual interactions, specifically in English and Nepali, catering to a broader user base.

The project involves the development of natural language processing (NLP) and natural language understanding (NLU) modules to facilitate accurate intent recognition and response generation. Together, these projects combine state-of-the-art machine learning techniques and advanced signal processing methods to create a robust and efficient conversational AI system. They not only demonstrate the practical application of AI and ML technologies but also pave the way for more intuitive and natural human-computer interactions.

3.3.2 System Analysis

3.3.2.1 Functional Requirements

Functional requirements are specifications that the client provides that describe the general functioning of the system, how the system responds to specific inputs, and how a system behaves inside a certain system. The use case diagram is used in this project to describe the general functioning of the system. Following are the functional requirements provided by the client for the project:

- Allow user to access the chatbot form the interface.
- Allow user to automatically talk with the bot.
- Allow user to chat through text or speech whichever is comfortable.
- Allow user to sequentially chat with the bot in both English and Nepali language.

The use case diagram below depicts the entire system's functioning:

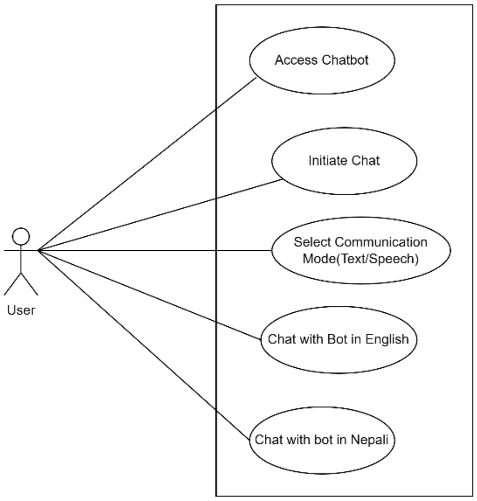


Figure 3.1 Use Case diagram of the Chatbot User Interface

Use Case Description

Here are some of the use case descriptions of the project:

Table 3.2 Use Case description to Access the Chatbot

Use Case ID	CHAT-01
Use Case Name	Chatbot user interface
Primary Actor	User
Secondary Actor	

Description	This use case enables the user to open the chatbot interface.
Pre-condition	The user has access to a device with the chatbot interface
Success Scenario	The chatbot interface is displayed, ready for interaction.
Failure Scenario	Interface is popped with error message.

Table 3.3 Use Case description of Initiate chat

Use Case ID	CHAT-02
Use Case Name	Initiate chat
Primary Actor	User
Secondary Actor	
Description	This use case enables the user to start a conversation with the chatbot.
Pre-condition	The user has accessed the chatbot interface.
Success Scenario	The chatbot initiates a greeting or prompts the user for input.
Failure Scenario	Interface is popped with error message.

Table 3.4 Use Case description for Select Communication Mode

Use Case ID	CHAT-03
Use Case Name	Select Communication Mode.
Primary Actor	User
Secondary Actor	
Description	This use case allows the user to choose their preferred communication method (text or speech).

Pre-condition	The user has initiated the chat
Success Scenario	The selected communication mode is activated.
Failure Scenario	Chatbot sends an error or sorry message.

Table 3.5 Use Case description for Chat with bot in Nepali or English

Use Case ID	CHAT-05
Use Case Name	Chat with Bot in Nepali or English
Primary Actor	User
Secondary Actor	
Description	It allows the user to communicate with the chatbot in Nepali or in English.
Pre-condition	The user has selected the communication mode and initiated the chat. The user is currently engaged in a chat session with the bot.
Success Scenario	The chatbot recognizes the new language and continues the conversation accordingly.
Failure Scenario	Chatbot sends an error or sorry message.

3.3.2.2 Non-functional Requirements

Non-functional needs are system functionalities that have a large influence on the system as a whole rather than on individual components.

The following are the project's fundamental non-functional requirements:

1. **Usability**: The system is simple to understand. Each component is used one after another. Simply overall functionality changes are based on change provided in the configuration file.

2. **Maintainability:** The system is easily maintainable in nature. Here, Object-oriented approach is used for defining the overall components of the system.

3.3.2.3 Feasibility Analysis

Operational Feasibility

The process of determining how well a proposed system addresses business issues or capitalizes on business opportunities is known as operational feasibility. The system is operationally feasible in nature as new data can be accessed through the application such that the business organization can expand to different countries.

Technical Feasibility

Technical feasibility is the process of determining an organization's or an individual's ability to build the overall planned system. Since, the overall project is done with the team, each member in team is technically feasible to implement the proposed project.

3.3.3 System Design

System design simply involves in organizing the working of the complex systems. System designer looks after different components of the systems, their interactions and provides detailed description of working of each component. This section shows overall design architecture of the system and some of the behaviors of the system. Overall performance of the system is completely dependent upon the design of the system.

Architecture design of system is the design of the overall structure, components, modules and subsystems of a system based on requirements provided.

3.3.3.1 Concept Diagram

Concept diagram is the visual representation which illustrates the relationships and connection between different elements of the systems. Concept diagram simply shows the outline of the overall architecture of the system.

Here is the concept diagram of Voice Activity Detector:

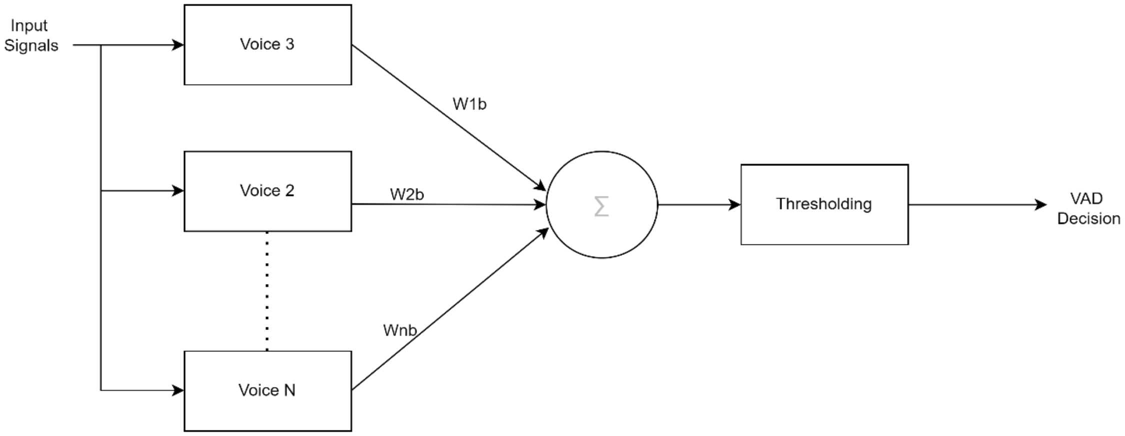


Figure 3.2 Concept diagram for VAD

The Voice-Activity-Detector (VAD) functions by processing incoming audio signals to determine whether they contain speech. The concept diagram for a VAD includes several key stages: input signals, input voices, summation of the voices, thresholding, and the final VAD decision. Initially, the system captures input signals, which are a mix of user speech and background noise. These input voices are then summed to form a comprehensive audio signal. The summed audio signal is analyzed and subjected to a thresholding process, where the system measures the signal's energy levels, frequency components, and temporal patterns. This thresholding process helps in distinguishing between speech and non-speech segments by comparing the analyzed characteristics against predefined thresholds. If the signal's characteristics surpass these thresholds, the VAD makes a positive decision, identifying the segment as containing speech. This decision allows the system to focus on relevant speech segments, filtering out non-speech parts, and thereby improving the efficiency and accuracy of the speech recognition process. Through these stages, the VAD ensures that only meaningful speech data is processed, enhancing the chatbot's ability to respond accurately and promptly.

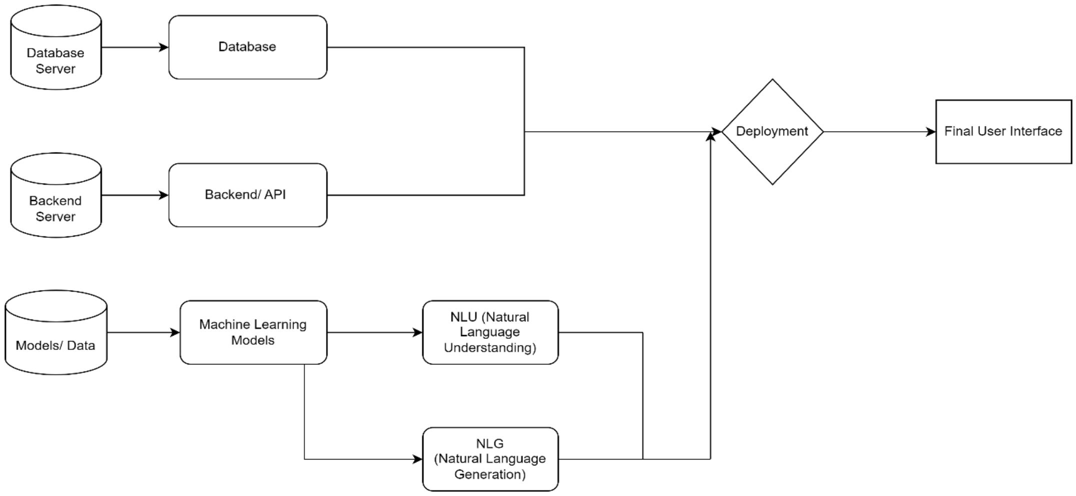


Figure 3.3 Workflow for Chatbot integration

Integrating a chatbot involves several key components working together seamlessly, including a backend server, database server, machine learning (ML) models for natural language processing (NLP) and natural language understanding (NLU), cloud deployment, and a user interface (UI). The backend server acts as the central hub, handling the main logic of the chatbot and orchestrating interactions between various components. It communicates with the database server to store and retrieve user data, conversation histories, and other relevant information. ML models for NLP and NLU are integrated to interpret and generate human-like responses, enabling the chatbot to understand user queries and provide appropriate replies. These models are often trained and fine-tuned to enhance accuracy and performance. Once the system is fully integrated, it is deployed on a cloud platform to ensure scalability, reliability, and easy access. Finally, the chatbot is accessible through a user-friendly UI, which can be embedded in various channels such as websites, mobile apps, or messaging platforms, allowing users to interact with the chatbot seamlessly. This integration ensures a robust and responsive chatbot that can handle complex interactions and provide valuable assistance to users.

3.3.3.2 Activity Diagram

Activity diagram is a form of diagram UML that indicate the sequence and flow of actions within a system. Here is the activity diagram of the text to speech and speech to text chatbots:

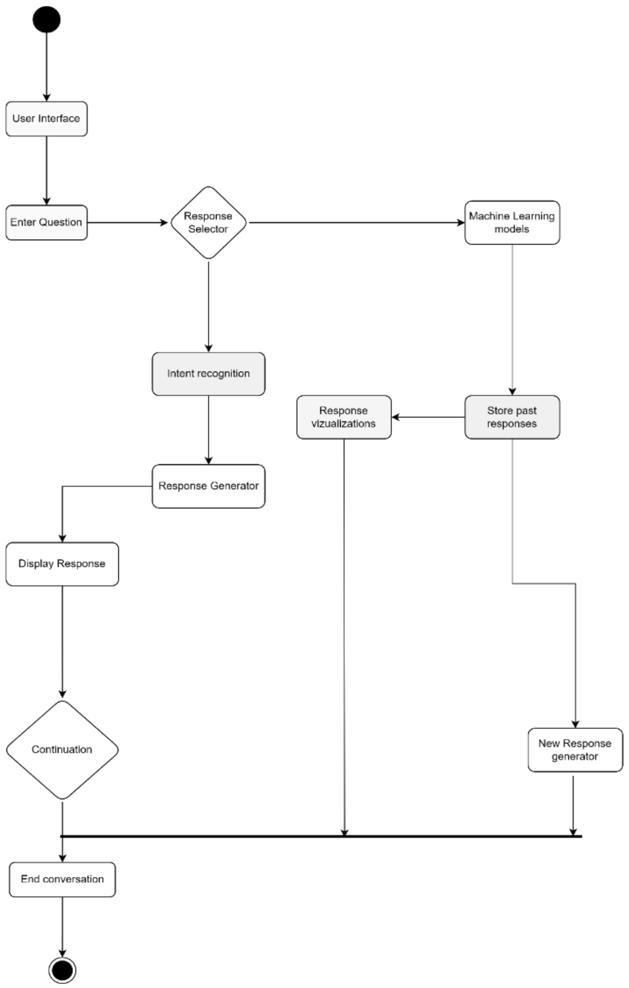


Figure 3.4 Activity diagram for a TTS Model Chatbot

The activity flow for the Text-to-Speech (TTS) model chatbot begins with user interaction through the user interface (UI), where the user inputs a question or query. Upon input, the system proceeds through several stages. Initially, it selects the appropriate response based on the nature of the input. Subsequently, machine learning models are employed for intent recognition, analyzing the input to understand the user's request accurately. Following this, the system visualizes potential responses relevant to the user's query. Based on the recognized intent and the visualized response options, the chatbot generates a new response, which is then displayed to the user through the UI. The user then decides whether to continue the conversation or end it based on the provided response, thereby determining the flow of the interaction. If the conversation continues, the process iterates back to the response generation stage; otherwise, the interaction concludes. This activity flow ensures the chatbot effectively interprets user queries, generates appropriate responses, and maintains a smooth and engaging conversation experience.

Here is the activity diagram for voice-activity-detector component of the project:

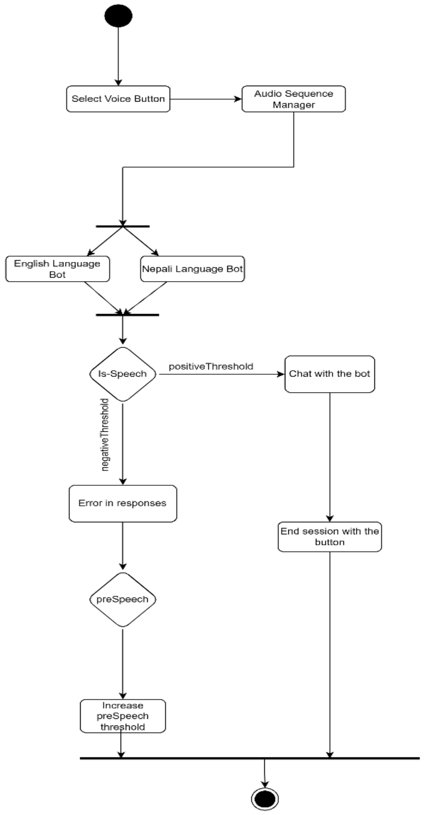


Figure 3.5 Activity diagram for voice-activity-detector

The provided activity diagram illustrates the organized sequence involved in managing audio sequences and facilitating user interactions with English and Nepali language chatbots. It begins with the user activating the voice input mode by selecting the "Select Voice" button on the user interface (UI). Following this, the system initiates the audio sequence manager to effectively handle incoming audio signals. Subsequently, the chatbot employs a language detection module to distinguish between English and Nepali speech. If the input is identified as speech, the chatbot engages in a conversation with the user, providing appropriate responses based on the detected language. In case of any processing or response generation errors, the system promptly alerts the user.

The interaction with the chatbot persists until the user opts to end the session by clicking the "End Session" button. Throughout the interaction, the system dynamically adjusts the preSpeech threshold to optimize speech recognition accuracy, ensuring a seamless user experience. In essence, the activity diagram provides a comprehensive visual representation of the systematic flow of user interactions and system responses in managing audio sequences and conducting conversations with English and Nepali language chatbots.

3.3.3.3 Sequence Diagram

Sequence diagram is the dynamic UML diagram which provides the interaction between two or more objects during a certain period of time. For each case in use case diagram, there are different functionalities such that each sequence diagram is for the specific use cases only.

The sequence diagram illustrates the coordinated flow of interactions among the Text-to-Speech (TTS), Speech-to-Text (STT), and Voice Activity Detector (VAD) components within the chatbot system. It begins with the user activating the voice input mode by selecting the "Select Voice" button on the user interface (UI). Upon activation, the system initializes the VAD to process incoming audio signals.

Here is the sequence diagram for the scrape website and APIs:

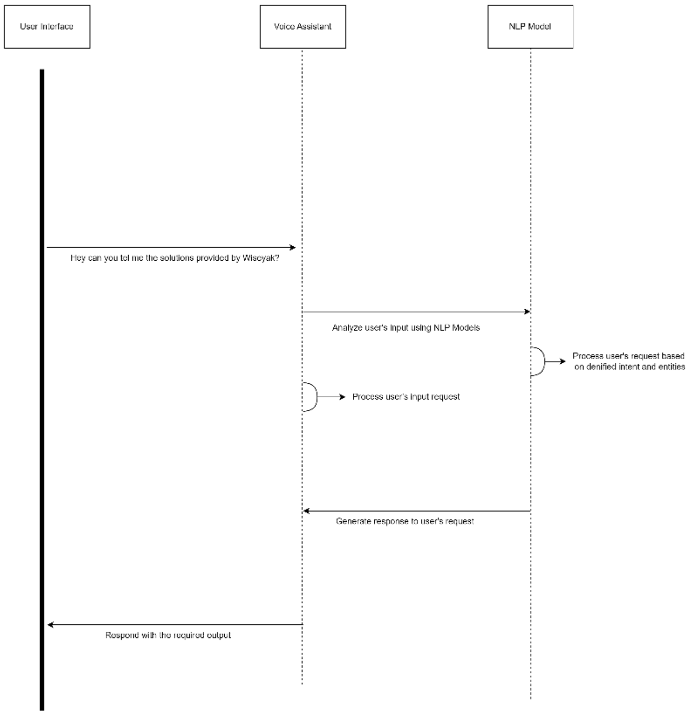

User Interface
Voice Assistant
NLP Model
Hey can you tel me the solutions provided by Wiseyak?
Analyze user's input using NLP Models
Process user's request based
on denified intent and entities
Process user's input request
Generate response to user's request
Respond with the required output

Figure 3.6 Sequence diagram for Scrape website and APIs

The VAD then analyzes the audio to identify speech segments, which are forwarded to the STT module for conversion into text. The text input is then passed to the chatbot's intent recognition module, which employs machine learning models to identify the user's intent. Based on the recognized intent, the system generates appropriate responses using the TTS module. These responses are displayed to the user through the UI.

Throughout the interaction, the VAD continues to monitor the audio signal, adjusting the preSpeech threshold dynamically to ensure accurate speech recognition. If any errors occur during the processing or generation of responses, the system promptly alerts the user. The conversation with the chatbot continues until the user decides to end the session by clicking the "End Session" button.

This sequence diagram provides a detailed representation of the synchronized flow of interactions among the TTS, STT, and VAD components, ensuring a seamless and efficient user experience in managing audio sequences and conducting conversations with both English and Nepali language chatbots.

3.3.4 Tools and Technologies Used

The tools that were used for the implementation and the design of the internship project are:

Python (v3.10.10)

Python is a general-purpose programming language for scripting. Python scripts in the project are used for writing the back API call which is used in the project.

Anaconda (v22.11.1)

Anaconda is Python-based cross-package management program created specifically for data science and related fields. Machine learning technologies such as Jupyter, Pandas, and Numpy are appropriately included during software installation. It also includes "conda," a command-line tool that allows users to manage packages and configure environments depending on the needs of the project.

Colab

Colab is an open-source, web-based tool for collaborative application assessment and development. It enables data visualization and discussion among team members to accelerate project development.

Git (v2.38.1)

Git is a software configuration management solution for coordinating distributed program versions. Through git, each feature change was logged, and team communication was correctly carried out.

Visual Studio Code

A portable, platform-neutral open-source IDE called Visual Studio Code is created to assist developers in building a range of apps. It offers a broad range of project development features.

Hugging Face Transformers

A library of open-source implementations of transformer models for variety of task.

LLM

LLM (Libraries, Languages, and Models) are essential components used in developing VAD, TTS, and STT functionalities. Python serves as the primary language, while various libraries and machine learning models are utilized for implementing specific tasks within each of these components.

3.4 Tasks/ Activities Performed

Here are some of the tasks/activities that were performed by the author in the project:

1. Voice Activity Detection (VAD) Project:
    - Conducted thorough research on existing VAD algorithms and techniques to gain a comprehensive understanding of the domain. Participated in brainstorming sessions with the team to discuss project requirements, goals, and implementation strategies.
    - Implemented VAD algorithms using Python and relevant libraries such as PyTorch, focusing on signal processing and machine learning techniques.
    - Collected and preprocessed audio data, including feature extraction, noise reduction, and segmentation, to prepare it for VAD model training. Developed machine learning models for voice activity detection, including data preprocessing, model training, evaluation, and optimization.
    - Conducted rigorous testing and evaluation of the VAD models using performance metrics such as accuracy, precision, recall, and F1-score. Collaborated with the team to integrate the VAD module into the larger chatbot system, ensuring seamless functionality and real-time performance.
2. Chatbot Project:
    - Contributed to the design and architecture of the chatbot system, incorporating VAD functionality for voice input processing.
    - Implemented the chatbot backend using the fastApi framework in Python, focusing on efficient HTTP request handling and response generation.

Integrated the VAD module into the chatbot system, allowing users to interact with the chatbot through both text and voice inputs.

- Developed natural language processing (NLP) and natural language understanding (NLU) modules for intent recognition and response generation using Python and relevant libraries.
- Collaborated with the team to define chatbot responses, including English and Nepali language support, and implemented multilingual capabilities. Conducted extensive testing and debugging of the chatbot system, ensuring robustness, reliability, and user-friendliness.

CHAPTER 4 - CONCLUSION AND LEARNING OUTCOMES

4.1 Conclusion

Concluding the three-month internship, the author found it to be a tremendously enriching experience, both personally and professionally. Throughout the internship period, the author has been deeply immersed in challenging projects focused on Voice Activity Detection (VAD) and Chatbot development. Engaging in various tasks and activities, including research, implementation, testing, and integration, has provided me with invaluable hands-on experience in artificial intelligence, machine learning, and natural language processing.

In the VAD project, my contributions involved extensive research, the implementation of VAD algorithms, and the seamless integration of these algorithms into the chatbot system. Similarly, my involvement in the Chatbot project allowed me to develop the backend infrastructure, implement NLP and NLU modules, and integrate multilingual capabilities, all of which contributed significantly to the project's success.

Furthermore, active participation in team meetings, code reviews, and feedback sessions facilitated collaboration and knowledge sharing with peers and mentors, which further enhanced my learning experience. This internship has not only equipped me with technical skills and knowledge but has also helped me develop essential soft skills such as problem-solving, teamwork, and communication.

The experience gained during this internship has laid a solid foundation for my future career in the field of artificial intelligence and machine learning. I'm deeply grateful for the opportunity provided by Wiseyak Solutions Pvt. Ltd. and am eager to apply the knowledge and skills acquired during this internship to future endeavors.

4.2 Learning Outcome

The internship period provided the following learning outcomes to the author. They are as follows:

4.2.1 Professional Level

- Practical experience in a real-world working project.
- Professional work behavior and attitude in the workplace.
- Team working skills, time management skills and boost to interpersonal communication.

4.2.2 Technical Level

- Proficient in Python programming language, including its libraries and frameworks, for implementing various AI/ML algorithms and applications.
- Hands-on experience in developing, training, and evaluating machine learning models for tasks such as Voice Activity Detection (VAD), Text-to-Speech (TTS), and Speech-to-Text (STT).
- Version control and team collaboration.
- Docker compose and route mapping
- Reporting tools and strategies.

References

van den Oord, A., Dieleman, S., Zen, H., et al. (2016). WaveNet: A Generative Model for Raw Audio. DeepMind.

Wang, Y., Skerry-Ryan, R., Jia, Y., et al. (2017). Tacotron: Towards End-to-End Speech Synthesis. Google Research.

Baevski, A., Zhou, Y., Mohamed, A., Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. Facebook AI Research.

Wang, Y., Skerry-Ryan, R., Jia, Y., et al. (2017). Tacotron: Towards End-to-End Speech Synthesis. Google Research.

Niu, C., Shan, H., & Wang, G. (2022). *SPICE: Semantic Pseudo-Labeling for Image Clustering.* Silicon Valley, California: arXiv.

Hannun, A., Case, C., Casper, J., et al. (2014). Deep Speech: Scaling up end-to-end speech recognition. Mozilla.

Baevski, A., Zhou, Y., Mohamed, A., Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. Facebook AI Research.

Weninger, F., Bergmann, C., & Schuller, B. (2015). Introducing CURRENNT: The Munich Open-Source CUDA RecurREnt Neural Network Toolkit. Journal of Machine Learning Research.

Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. Advances in Neural Information Processing Systems (NeurIPS).

Reynolds, D. A., & Rose, R. C. (1995). Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. IEEE Transactions on Speech and Audio Processing. Wu, Z., & Falk, T. H. (2010).

Speech Quality Evaluation for Hearing Aids Using Gaussian Mixture Models. IEEE Transactions on Audio, Speech, and Language Processing. Parada, C., Rodriguez, R., & Metze, F. (2011).

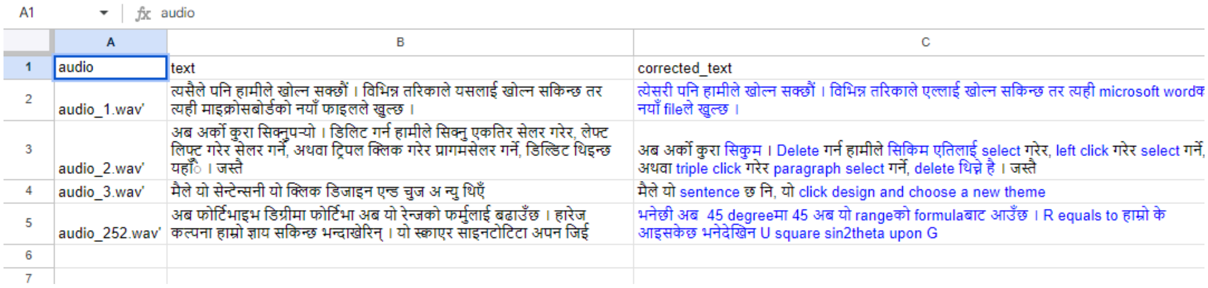
Improved Voice Activity Detection Using Long-Term Signal Variability. INTERSPEECH.Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). *Show and Tell: A Neural Image Caption Generator.* California: Google.

APPENDIX

APPENDIX:

Screenshots

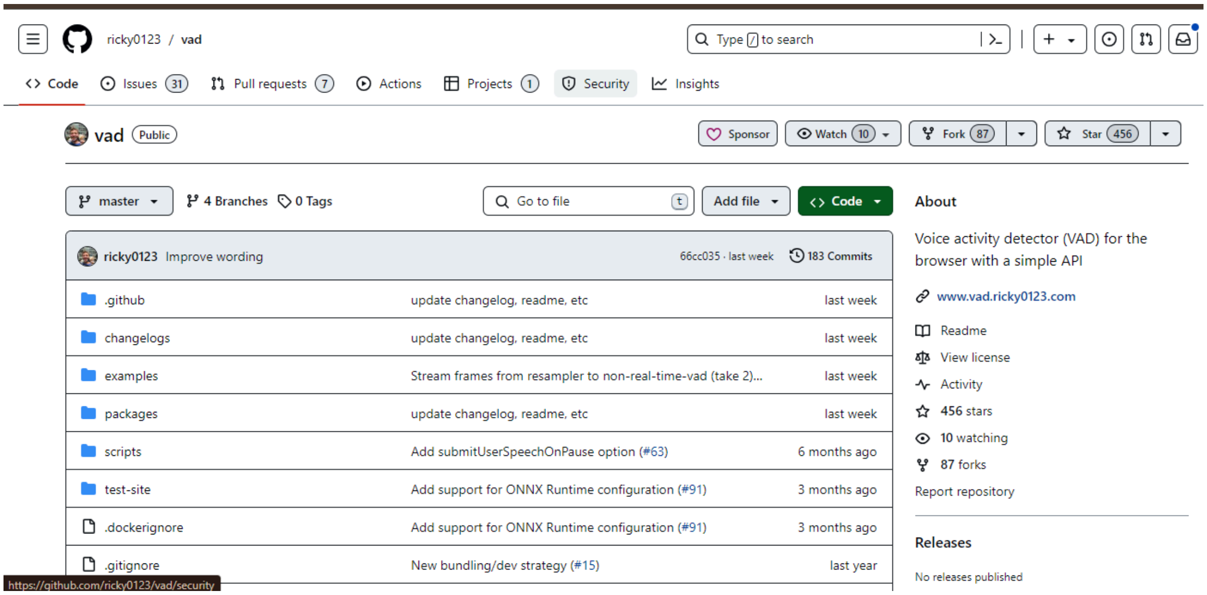
Data Collection and preparation



A1 ▾ | fx audio

	A	B	C
1	audio	text	corrected_text
2	audio_1.wav'	त्यसैले पनि हामीले खोल्न सक्छौं । विभिन्न तरिकाले यसलाई खोल्न सकिन्छ तर त्यही माइक्रोसबोर्डको नयाँ फाइलले खुल्छ ।	त्येसरी पनि हामीले खोल्न सक्छौं । विभिन्न तरिकाले एल्लाई खोल्न सकिन्छ तर त्यही microsoft wordक नयाँ fileले खुल्छ ।
3	audio_2.wav'	अब अर्को कुरा सिक्नुपऱ्यो । डिलिट गर्न हामीले सिक्नु एकतिर सेलर गरेर, लेफ्ट लिफ्ट गरेर सेलर गर्ने, अथवा ट्रिपल क्लिक गरेर प्रागमसेलर गर्ने, डिल्डिट थिइन्छ यहाँ ा । जस्तै	अब अर्को कुरा सिकुम । Delete गर्न हामीले सिकिम एतिलाई select गरेर, left click गरेर select गर्ने, अथवा triple click गरेर paragraph select गर्ने, delete थिन्ने है । जस्तै
4	audio_3.wav'	मैले यो सेन्टेन्सनी यो क्लिक डिजाइन एन्ड चुज अ न्यु थिएँ	मैले यो sentence छ नि, यो click design and choose a new theme
5	audio_252.wav'	अब फोर्टिभाइभ डिग्रीमा फोर्टिभा अब यो रेन्जको फर्मुलाई बढाउँछ । हारेज कल्पना हाम्रो ज्ञाय सकिन्छ भन्दाखेरिन् । यो स्काएर साइनटोटिटा अपन जिई	भनेछी अब 45 degreeमा 45 अब यो rangeको formulaबाट आउँछ । R equals to हाम्रो के आइसकेछ भनेदेखिन U square sin2theta upon G
6			
7			

V-A-D Repository



ricky0123 Improve wording 66cc035 · last week 183 Commits

.github	update changelog, readme, etc	last week
changelogs	update changelog, readme, etc	last week
examples	Stream frames from resampler to non-real-time-vad (take 2)...	last week
packages	update changelog, readme, etc	last week
scripts	Add submitUserSpeechOnPause option (#63)	6 months ago
test-site	Add support for ONNX Runtime configuration (#91)	3 months ago
.dockerignore	Add support for ONNX Runtime configuration (#91)	3 months ago
.gitignore	New bundling/dev strategy (#15)	last year

About

Voice activity detector (VAD) for the browser with a simple API

www.vad.ricky0123.com

Readme

View license

Activity

456 stars

10 watching

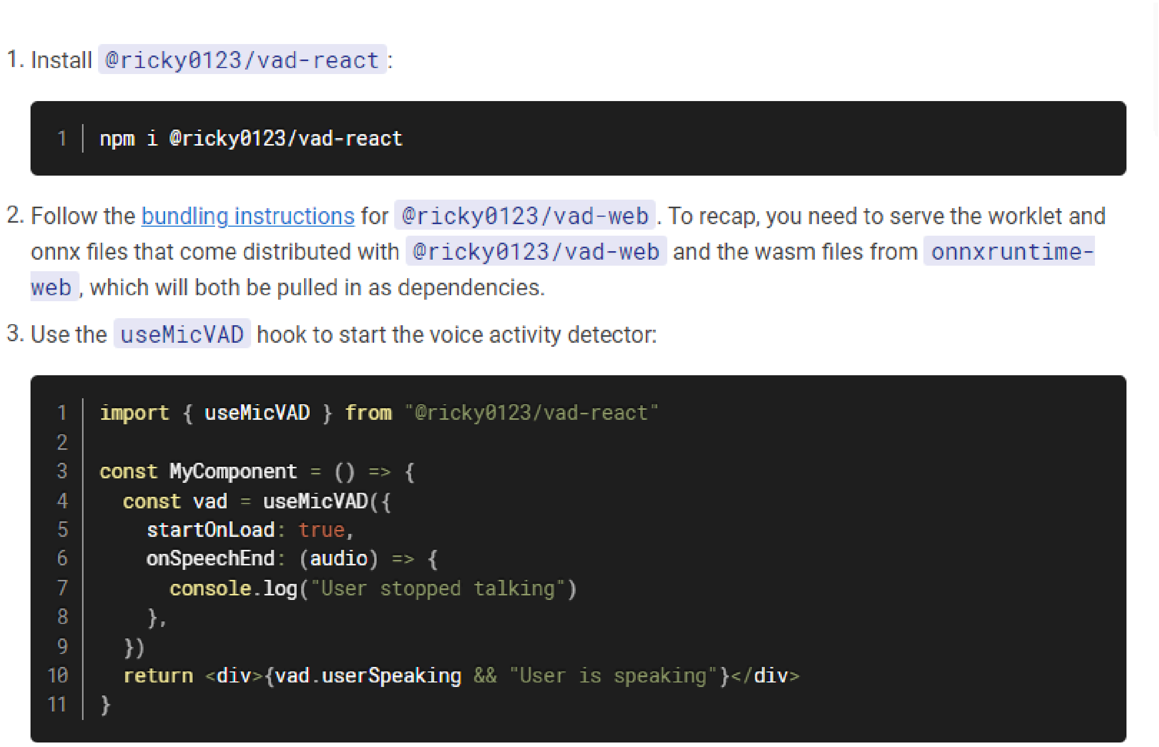
87 forks

Report repository

Releases

No releases published

Integration in the app using React Framework.



1. Install `@ricky0123/vad-react`:

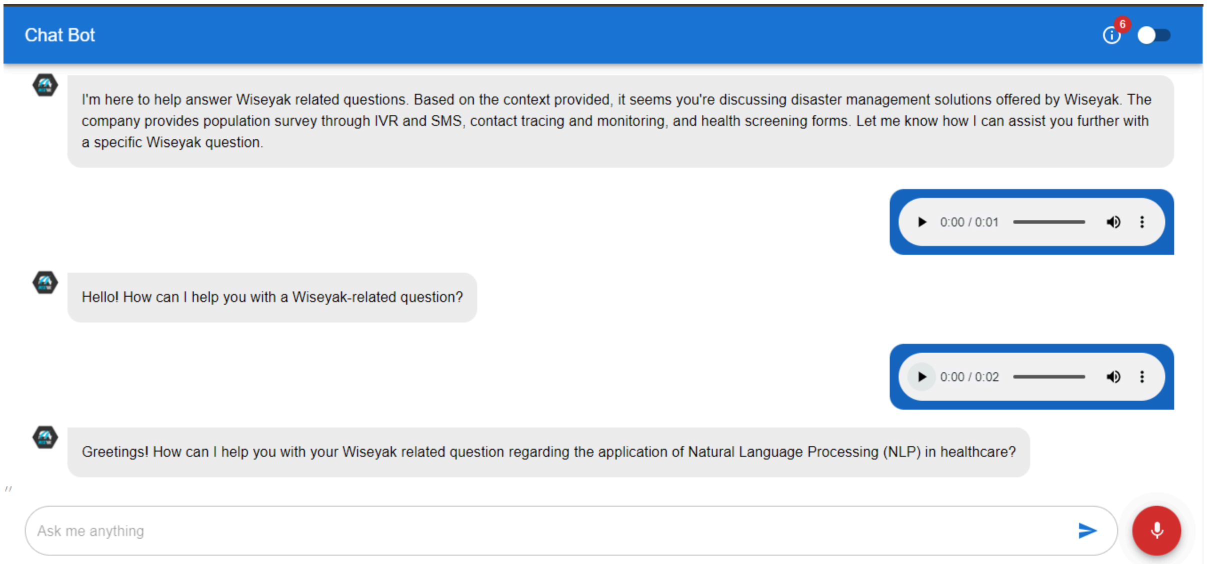
```
npm i @ricky0123/vad-react
```

2. Follow the bundling instructions for `@ricky0123/vad-web`. To recap, you need to serve the worklet and onnx files that come distributed with `@ricky0123/vad-web` and the wasm files from `onnxruntime-web`, which will both be pulled in as dependencies.
3. Use the `useMicVAD` hook to start the voice activity detector:

```
import { useMicVAD } from "@ricky0123/vad-react"

const MyComponent = () => {
  const vad = useMicVAD({
    startOnLoad: true,
    onSpeechEnd: (audio) => {
      console.log("User stopped talking")
    },
  })
  return <div>{vad.userSpeaking && "User is speaking"}</div>
}
```

Text to speech and speech to text Chatbot

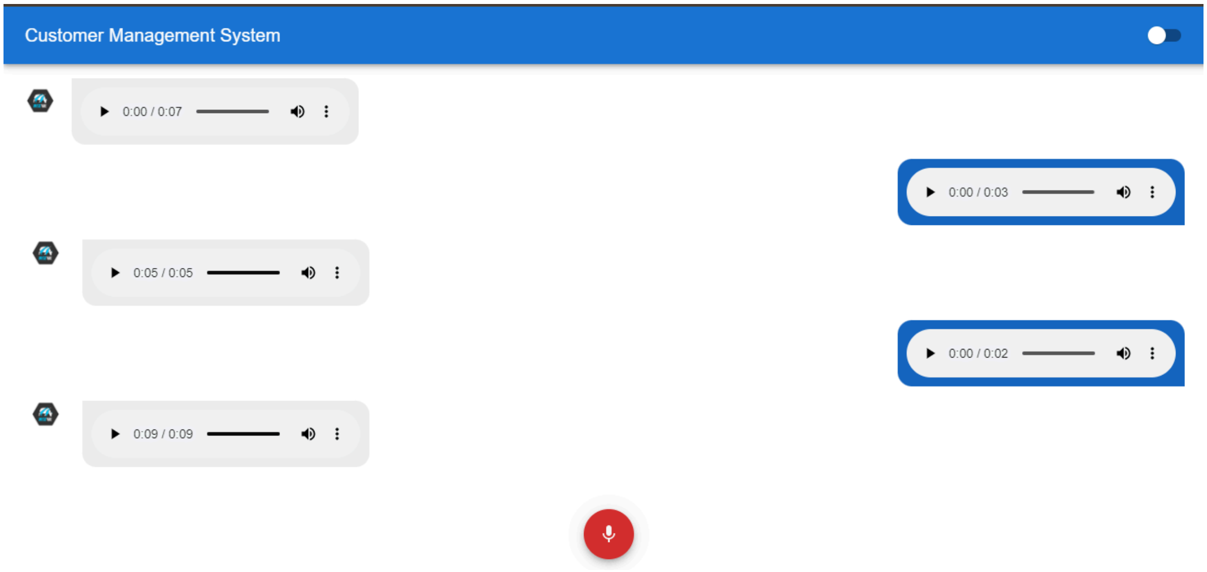


I'm here to help answer Wiseyak related questions. Based on the context provided, it seems you're discussing disaster management solutions offered by Wiseyak. The company provides population survey through IVR and SMS, contact tracing and monitoring, and health screening forms. Let me know how I can assist you further with a specific Wiseyak question.

Hello! How can I help you with a Wiseyak-related question?

Greetings! How can I help you with your Wiseyak related question regarding the application of Natural Language Processing (NLP) in healthcare?

Speech only Chatbot


Customer Management System
0:00 / 0:07
0:00 / 0:03
0:05 / 0:05
0:00 / 0:02
0:09 / 0:09