Ezhil Raj

Modakurichi,
Erode,
Tamil Nadu,
India.
Tel +91 6380688699

ezhil6999@gmail.com

GitHub: Ezhil6999
LinkedIn: Ezhil raj. S
Profile: Ezhil raj
Leetcode: Ezhil6999

ABOUT

I am Ezhil Raj, an AI Engineer at SynBrains Solution with 2+ years of professional experience in the field of Natural Language Processing and Gen AI. My expertise encompasses working with diverse data types, including text, image, and audio, to develop innovative solutions and advance NLP technology.

SKILLS

- Python
- Java
- MySQL
- Redis
- Pytorch
- Machine Learning
- Deep Learning
- Agentic AI
- AWS
- GitHub
- Docker
- MongoDB

TRAINING & CERTIFICATIONS

- In 2022, I gained valuable experience in SEO (Search Engine Optimization) through my role as a content writer at Tech Perl.
- In the Value-Add Course on Java, I received both an advanced and creative level appreciation Certificate for my work. This course equipped me with advanced skills in Java programming, including object-oriented concepts, data structures, and algorithms.
- Successfully completed the 'C' Certificate in the National Cadet Corps (NCC).

Experience

Company Name :Vectone Mobiles solution — June 2023 - December 2024

➔ I have 1 year 5 months of experience worked in an NLP team, contributing to multiple projects for the growth of the company.

Company Name :Synbrains Solution — April 2025 - Currently

➔ I have 7 months of experience working in GenAI related projects.

PROJECTS

SynBrains Solution — April 2025 - Currently

Agentic and MCP Projects:

➔ I have worked on various intelligent agent projects, including table booking agents and sales agents, implemented using both N8N workflows and custom code-based architectures. For Agentic AI, I have extensively used the OpenAI Agent library to build and orchestrate autonomous agents capable of complex decision-making and data-driven interactions.

➔ Most of my projects involved creating chatbots integrated with MongoDB and MySQL databases, enabling dynamic data retrieval and conversational intelligence.

➔ In addition, I have practical experience with Fast MCP (Model Context Protocol), using it to connect and manage context-aware models efficiently.

➔ My work reflects strong expertise in building scalable agentic systems, LLM integration using OpenAI APIs, and designing automation workflows that bridge AI with real-world data.

Vectone Mobile solution

Integration: June 2023 - April 2024

- ➔ I have extensive experience in integrating Large Language Models (LLMs), Text-to-Speech (TTS) systems, Speech-to-Text (STT) technologies, and other essential components.

Vectone Mobile solution

Classification Problems: January 2024 - March 2024

- ➔ I have worked on both image and text classification tasks. The text classification was focused on intent recognition, while the image classification involved detecting the presence of a helicoil. These tasks were part of a robotics project. It is a binary classification problem using a pretrained res-net model.

Vectone Mobile solution

E-commerce Website: August 2024 - December 2024

- ➔ In our project, we employed Playwright and Beautiful Soup to perform manual web scraping across multiple websites. The extracted data underwent necessary preprocessing before being stored both locally and in a MySQL database.

- ➔ Based on user queries, we retrieve the relevant data from the locally stored database. Additionally, we have trained a model to generate taxonomy aligned with the context of the user query, enhancing the relevance and structure of the responses.

Vectone Mobile solution June 2024 - August2024

Smart Speaker:

- ➔ In the smart speaker, we need to integrate text-to-speech, speech-to-text, and large language models (LLMs) for conversational responses. This integration has already been accomplished using our previous experience and expertise.
- ➔ Currently, I am addressing challenges related to wake word detection. Specifically, I am in the process of training a wake word model utilizing Long Short-Term Memory (LSTM) networks.

Vectone Mobile solution May 2024 - June 2024

Speech Diarization and Speaker Identification:

- ➔ In speech diarization, we identify the number of speakers present in the audio. We then segment the audio according to speaker turns and provide a corresponding transcription with timing information.
- ➔ In this implementation, we utilized open-source packages and models, which provided good accuracy.
- ➔ In our speaker identification system, we utilized an open-source model. We assess audio similarity using embeddings, which enables identification in less than 50 milliseconds. For optimal accuracy, audio with a medium pitch is preferred. Our testing, conducted with 20-30 speakers, has demonstrated improved accuracy.
- ➔ Speech diarization is performed using Redis.
- ➔ Speaker identification was incorporated into the Speech-to-Text system.

Vectone Mobile solution March 2024 - May 2024

Text to Speech:

- ➔ We have implemented an open-source text-to-speech model, utilizing both TCP and UDP sockets for data transmission.
- ➔ TCP serves as the primary socket for communication, while UDP is utilized for streaming audio bytes (RTP packets). We continuously monitor the TCP sockets. If a session needs to be terminated, a message is sent via TCP, prompting us to stop the TTS playback.
- ➔ The open-source multilingual TTS model proved to be unstable, so we implemented a more reliable open-source TTS model trained on a single language.
- ➔ The implementation of speech-to-text technology facilitates the seamless development of text-to-speech systems.

Vectone Mobile solution January 2024 - March2024

ChatBot:

- ➔ We have developed a chatbot that integrates an open-source model with advanced retrieval methods.
- ➔ We have implemented a feature that stores individual session memories of previous user interactions. The primary goal is to accurately retrieve the relevant context from sources such as PDFs, CSV files, or web

scraping.
- ➔ In this implementation, we utilized Redis to stream the LLM conversation.
- ➔ The ChatBot includes specialized implementations beyond just answering questions. We've developed a system to facilitate appointment bookings with a doctor or another individual. When a user requests to book an appointment, we trigger a Named Entity Recognition (NER) model to extract key details such as the name, meeting time, and other relevant information.
- ➔ We attempted to fine-tune the open-source LLM model, and while the results were satisfactory, they did not fully meet our expectations.
- ➔ The booking process has finally been completed using the prompt.

Vectone Mobile solution October 2023 - January 2024

Virtual Trial Room:

- ➔ This project has been highly valuable for me, enhancing my knowledge in image processing and providing a solid foundation in model training.
- ➔ The primary objective of this project is to accurately merge two images: one featuring clothing and the other depicting a human figure. The model aims to identify the dressing area on the human image and seamlessly replace the existing attire with the clothing from the other image, ensuring it appears as a naturally wearable outfit on the person.
- ➔ The process involves several preprocessing stages, including keypoint extraction, human parsing, in-painting using keypoints and human parsing, masking the human image, and masking the clothing image.
- ➔ We have implemented various open-source models such as HD-VTON, LADI-VTON, OOT Diffusion, and Stable VTON, but none have met our expectations. Consequently, we decided to develop our own solution.
- ➔ Due to resource constraints, we are unable to proceed with the project.

Vectone Mobile solution August 2023 - October 2023

RAG Framework :

- ➔ This project has been a significant achievement for me, given my extensive individual contributions and the substantial role I played in its development.
- ➔ We have developed a bespoke implementation of the RAG (Retrieval-Augmented Generation) architecture, encompassing document loading, text segmentation, embedding, storage, cosine similarity computation, and retrieval processes.
- ➔ In the RAG framework, we unlock a range of advanced capabilities, including ensemble methods, RAG fusion, the combination of ensemble with RAG fusion, and named entity recognition retrieval. These features are achieved through the strategic use of prompt engineering and retrieval techniques.these techniques are used to make the output or context meaningful for the understanding of the model.
- ➔ langchain and llama index used openAI for the users answer.Unlike, long chain and llama index we have used open source LLM model for the users answer.
- ➔ We have completed all necessary preparations, including finalizing the code for GitHub and making it ready for deployment on PyPI (Python Package Index).

Vectone Mobile solution June 2023 - August 2023

Speech to Text :

- ➔ In my speech-to-text project, I utilized TCP sockets and multithreading to enable concurrent processing. I developed WebSocket connection code to capture input and filter out non-human audio, ensuring only relevant speech bytes were processed. The speech-to-text model runs independently within an API. Upon receiving the correct audio bytes, the system calls this API to transcribe the speech, then returns the resulting text via WebSocket.
- ➔ I have been involved in projects where I built API code for models, WebSocket integration, and multithreading, and contributed to every aspect of the speech-to-text process.
- ➔ We successfully implemented the solution using an open-source model that delivers high accuracy and fast inference, with a response time of less than 500 milliseconds.
- ➔ We have integrated our speech-to-text service with DNS-based Auto-Scaling to enhance its performance and reliability.

EXTRA CURRICULAR ACTIVITIES

I actively participated in the National Cadet Corps (NCC) from 2020 to 2023, and I was honored to achieve the rank of Sergeant. During my tenure, I successfully completed both the 'B' Certificate Exam and the 'C'

Certificate Exam, earning a B grade in both examinations. These exams assessed my leadership skills, discipline, physical fitness, and knowledge of military subjects. Through my NCC involvement, I developed qualities such as teamwork, self-discipline, and a strong sense of responsibility.

LANGUAGES

- Tamil ●●●●
- English ●●●●

EDUCATION

Degree	Institution	Year of Passing	Percentage
B.Tech (Information Technology)	Velalar College of Engineering and Technology, Thindal, Erode, Tamil Nadu.	2023	84.5

DECLARATION

I hereby declare that the information provided in this resume is true, complete, and accurate to the best of my knowledge. I also consent to the collection, storage, and processing of my personal data for recruitment purposes. I am confident in my abilities and look forward to contributing my skills and expertise to the growth and success of the organization.

Ezhil raj

Erode,

November 13, 2025

Ezhil raj.