DINESH KUMAR T

AI / ML Engineer | NLP, LLMs & RAG | Speech-to-Text & Voice AI

Chennai, India | +91 7904604102 | dineshtamilselvan.24@gmail.com | linkedin.com/in/dinesh-kumar-tamilselvan

SUMMARY

AI/ML Engineer with 3 years of experience building production NLP, speech and voice AI systems for contact centres. I've shipped an intent classification API that replaced Rasa NLU (about 15% better accuracy), built real-time speech-to-text pipelines, and worked on LLM and RAG applications using LangChain, Hugging Face and vector databases. Comfortable taking a model from experiment to a Dockerised API on AWS, and debugging it when it misbehaves in production.

SKILLS

ML & Deep Learning: PyTorch, TensorFlow, scikit-learn, Hugging Face Transformers, BERT, Sentence-BERT, TF-IDF, Logistic Regression, KNN, Naive Bayes, Random Forest, SVM, cross-validation, F1 / precision / recall
LLMs & GenAI: OpenAI GPT APIs, open-source LLMs, LangChain, RAG, embeddings, semantic search, vector search on Amazon Aurora PostgreSQL (pgvector), prompt engineering
Speech & Voice AI: Whisper, NVIDIA Parakeet-TDT, Vosk, Kaldi, VAD, speaker diarization, audio preprocessing (FFmpeg), Asterisk, SIP, IVR, Python AGI
Backend & MLOps: Python, FastAPI, Flask, REST APIs, WebSockets, Kafka, Redis, Docker, Kubernetes, GPU inference, model serving, monitoring, latency and concurrency tuning
Cloud & Data: AWS (EC2, Lambda, S3, Aurora PostgreSQL), SQL / MySQL / PostgreSQL, Pandas, NumPy, ETL pipelines, Power BI

EXPERIENCE

Data Scientist / AI Engineer — Pulse Telesystems, Chennai

Mar 2024 – Present

NLP & intent classification

- Designed and deployed a two-stage intent classification API (process first, then intent) that replaced Rasa NLU and improved intent accuracy by about 15%.
- Combined rule-based logic with ML and deep learning models (TF-IDF, Logistic Regression, KNN, BERT) and compared them using weighted F1 and cross-validation before each release.
- Used Sentence-BERT embeddings from Hugging Face for semantic similarity and for matching new intents without retraining.
- Cut memory use and inference latency by loading per-process models on demand; served predictions through REST APIs (Flask, FastAPI) with confidence scores, deployed on AWS EC2, Lambda and S3.

LLMs & RAG

- Built LLM and RAG pipelines with LangChain, using OpenAI GPT APIs and open-source LLMs, with document embeddings stored in Amazon Aurora PostgreSQL (pgvector) for retrieval.

Voice AI & telephony

- Developed Python AGI services on Asterisk for DTMF handling, call routing and IVR flows used by live voice bots.
- Traced SIP signalling, call flows and MixMonitor recordings to find root causes of production incidents such as routing failures, queue timeouts and transfer issues.

Data engineering & analytics

- Built ETL pipelines from MySQL and AWS S3 into Python services, reducing processing time by about 20%.
- Built Power BI dashboards and tuned SQL queries for the operations team, contributing to about 10% better operational efficiency.

PROJECTS

Real-Time Speech-to-Text & Audio Processing Platform

Whisper, Parakeet-TDT, Vosk, Kaldi, FastAPI, Kafka, Redis, WebSockets, Docker

- Built and benchmarked STT pipelines across four engines for real-time and post-call transcription, comparing accuracy, latency and resource use.
- Designed a streaming architecture with FastAPI, Kafka, Redis and WebSockets for concurrent audio processing, with VAD and FFmpeg preprocessing before inference.
- Added speaker diarization and transcript segmentation to produce speaker-labelled transcripts.
- Added GPU inference, request limiting, concurrency controls and memory-based restarts; containerised with Docker and added monitoring for stable operation under load.

Model Validation & Risk Analytics

- Ran performance validation, regression and stress testing on ML models and presented risk findings to business stakeholders.

EDUCATION

PG Program in Data Science & Analytics — Imarticus Learning, Chennai (Jun 2023 – Jan 2024)
B.E. Mechanical Engineering — Anna University, Chennai (Jun 2014 – May 2018)

CERTIFICATIONS

Python for Data Science and Machine Learning Bootcamp (Udemy) | HackerRank: Python, SQL, Problem Solving