LYLE OKOTH

MACHINE LEARNING ENGINEER | SPEECH AND AUDIO PROCESSING

LinkedIn | Portfolio | GitHub | Blog | Huggingface
+254797668066 | okothlyleochieng@gmail.com
Nairobi, Kenya

CAREER OBJECTIVE

A professional Machine Learning Engineer with 3+ years of experience designing, training, building and deploying small, fast and highly expressive multilingual speech synthesis, speech translation and speech transcription models that empower content creation, remove language barriers and enable seamless interaction with computers. I mainly focus on Kenyan languages and work on data collection and annotation, data modelling, model training and deployment as well as ui design for speech based applications.

SKILLS AND CORE COMPETENCIES

- Speech Data annotation
- LLM Training and Fine-Tuning
- Data Modelling
- Web Scraping
- Writing research papers
- Retrieval Augmented Generation
- Language Modelling
- MlOps and LLMOps
- Conversation Design
- Synthetic data generation
- System Design
- Multi-Agent workflows
- Technical writing
- LLM Evaluation
- GPU Programming

TECHNICAL SKILLS

Languages: *Python, C++, JavaScript, SQL*
Frameworks and Libraries: *Flask, FastAPI, Django, React, NumPy, Pandas, Matplotlib, Scrapy, Pytorch, Scikit-learn, Transformers, langchain, spacy,, crew-ai, comfy-ui, llamaindex, huggingface, NLTK, LangGraph, Triton, Cuda*
Cloud: *AWS, GCP*
Databases and Storage: *PostgreSQL, Elasticsearch, Redis*
Others: *Apache Kafka, Apache Spark, Docker, Kubernetes, Celery*

EXPERIENCE

Machine Learning Engineer – Sautiflow labs **August 2023 – Present**

- Built an audio data collection, editing and annotation platform that leverages finetuned audio models resulting in faster and cheaper data labelling
- Collected and annotated high quality speech datasets for training translation, transcription and synthesis models resulting in small and highly accurate models
- Craft data pipelines for synthetic data generation that generated diverse data sets that increased model generalization and handled edge cases effectively
- Trained and fine-tuned small and fast multi-lingual speech models that were cheaper to deploy, significantly reduced time to first token and enabled realtime use.
- Setup rigorous logging and tracing systems for monitoring token usage and detecting model drift that enabled us to further optimize the models for speed and accuracy
- Setup benchmarks for testing the effectiveness of model experiments
- Setup reinforcement learning training pipelines for preference tuning and guarding against misuse of models including prompt injections resulting in better overall user experience
- Read and implemented the latest research on audio and speech models resulting in models that are very close to the state of art models
- Worked with software engineers to design easy to use web based interfaces for speech based products including real time streaming.

Software Engineer – Sautiflow labs **January 2022 – Present**

- Built a fast and highly scalable synchronous and asynchronous backend for realtime and batch processing of user requests. This enabled both realtime streaming for speech generation and translation as well as batch inference.
- Designed and built a highly intuitive user interface for working with speech based products. This includes real time translation, transcription as well as synthesis, resulting in high retention rates.
- Built an onboarding workflow that leverages AI to take new users through the usage of our products resu;lting in very high adoption rates
- Built a data studio for data collection and annotation that leveraged collaboration between humans and AI to increase the speed and quality of data labelling

- Built robust security measures for authentication and authorization to prevent API misuse by bots ensuring an overall greater user experience for real users
- Setup rigorous logging and tracing systems for monitoring APi usage to prevent downtimes, API misuse which improved overall user experience
- Wrote detailed documentation that made on-boarding new hires easier

ACADEMIC QUALIFICATIONS

- **Bsc. Computer Security and Forensics -** Jaramogi Oginga Odinga University of of Science and Technology
- **High School Diploma –** Friends School Kamusinga – (81/84)

LEADERSHIP AND VOLUNTEER CAUSES

- **OMDENA - Volunteer** **2024**
- **Co-Founder –** SautiFlow Labs **2022**

PERSONAL PROJECTS

Savanna Faces GitHub | Demo

- A web application for generating realistic looking Kenyan faces. Uses a diffusion model trained on a custom dataset.
- Includes payment, logging and tracing, user management, authentication and authorization. Deployed on render

SnapTube GitHub | Demo

- A web application that extracts timestamps from YouTube videos using a custom trained small language model. Built to be used to teach beginners how to pretrain and finetune a large language model. The data collection pipeline is also included
- The model is pre-trained and finetuned then use reinforcement learning for alignment
- Includes payment, logging and tracing, user management, authentication and authorization as well as admin dashboard. Deployed on render.

Savanna Faces GitHub | Demo

- A web application for generating realistic looking Kenyan faces. Uses a diffusion model trained on a custom dataset.
- Includes payment, logging and tracing, user management, authentication and authorization. Deployed on render

REFERENCES

AVAILABLE ON REQUEST