Tanmay Khandelwal

+1 (201) 268-9949 tk3309@nyu.edu /tanmayy24 /tanmayy24 Google Scholar Website

Education

New York University (NYU), Courant Institute of Mathematical Sciences **Sep 2023 - May 2025**

Master of Science in Computer Science (Concentration in AI) ***GPA:*** *4.0/4.0* *New York, USA*

Relevant Courses: Deep Learning *(Prof. Yann LeCun)*, Natural Language Processing, Machine Learning, Real-time and Big Data Analytics, Computer Vision, Advanced Data Structures and Algorithms, Programming Languages, Operating Systems

Birla Institute of Technology and Science (BITS), Pilani **Aug 2017 - Jul 2021**

B.E. (Hons) in Electrical and Electronics Engineering ***GPA:*** *8.43/10.0 (First Division)* *Pilani, India*

Relevant Courses: Information Retrieval, Data Mining, Object Oriented Programming (OOP), Neural Networks and Fuzzy Logic

Experience

Amazon **June 2024 - Aug 2024**

Applied Scientist Intern *Sunnyvale, California*

- Developed a **transformer-based two-tower network** for Amazon Music using **LoRA adaptor** for **parameter-efficient fine-tuning (PEFT)**, **attention-based feature fusion**, **mixture of experts (MoE)**, and distributed training for **embedding-based retrieval (EBR) on large scale dataset**; Improved real-time search and recommendation, increasing accuracy by **61.5%** on low-performing queries.
- Optimized search ranking by implementing a **popularity-weighted contrastive loss**, achieving a **25.17%** increase in **Recall@20** while utilizing **generative AI** for **synthetic data generation** and **user search logs** to further enhance performance.
- Deployed **FAISS-based** production system on **AWS** with **guardrails**, incorporated **LLMs via Amazon Bedrock** to leverage **prompt engineering** for generating and indexing key phrases for **lyrics queries**, improving accuracy and achieving **<80ms** response time.

Music and Audio Research Laboratory (MARL), NYU **Sep 2023 - Present**

Graduate Research Assistant *New York, USA*

- Developing **diffusion-based text-to-audio generation model** integrating **contrastive language-audio pretraining (CLAP)** with a **variational autoencoder (VAE)** decoder and **generative adversarial network (GAN)** vocoder.
- Contributed to open-source **Python** library: **Soundata** [Source] [Paper] with **32k downloads**, introducing data loaders, interactive visualization tools, and standardized usage for enhanced reproducibility; published in **journal of open source software (JOSS)**.

Fortemedia Inc. **Sep 2021 - Jul 2023**

Machine Learning Engineer *Singapore*

- Engineered **low-complexity** acoustic event detection system [Paper] for **speech recognition** embedded devices, integrating **attention modules** and **Bi-GRU**, resulting in a **34.1%** improvement in PSDS metrics while reducing model complexity by **27.6%**.
- Devised **multi-task learning (MTL)** framework [Paper] with two-stage **semi-supervised learning (SSL)** system for **speech-to-text software**, employing **transformer** and **conformer** to model global and local sequences; Improved PSDS by **45.5%**.
- Developed and **deployed** a scalable, **real-time** infant cry detection system [Paper] utilizing **depthwise-separable convolutions** and **Django REST APIs** for backend services; Accomplished an F-score of **0.738** on the curated dataset.

Bajaj Finserv Health Limited **Jul 2020 - Aug 2021**

Software Developer Intern *Pune, India*

- Engineered personalized medication **recommendation system** with past history and patterns-based quick suggestions using **Elasticsearch** and **SQL**, reducing E-consult time with **response time** $<$ **50ms**.
- Pioneered highly scalable **microservices** for a doctor's practice management system [Link] in **SpringBoot** and **NodeJs**, combined with **MongoDB** and **Redis**, using **Docker** and **Kubernetes** for **1000+ daily** E-consult users.
- Generated automated data scraping using **Selenium**, integrated third-party services, implemented **CI/CD system** with **Azure DevOps**, conducted **unit testing**, minimized code duplicity and set up **Elasticsearch, Logstash, and Kibana** stack monitoring.

Publications

- "A Multi-Task Learning Framework for Sound Event Detection using High-level Acoustic Features", ***INTERSPEECH*** [DOI]
- "Leveraging Audio-Tagging Assisted Sound Event Detection using Weak Generated Labels and Frequency Dynamic Convolutions (FDY-CRNN)", ***IEEE SSP*** [DOI]

Technical Skills

Programming Languages: Python, C++, Java, JavaScript, Typescript, Matlab, Scala, ML
Tools: GitHub, Apache Kafka, Spark, Hadoop, Kubernetes, Docker, Git, GPU, Nvidia CUDA, MySQL, Agile
Frameworks: PyTorch, TensorFlow, PyTorch Lightning, Keras, OpenCV, Hugging Face, SciKit-learn, Matplotlib/Seaborn, Numpy, Pandas, NLTK, spaCy, XGBoost, Kubeflow, Prometheus, Grafana, Microsoft Azure, AWS (Sagemaker, Lambda, EC2), Google Cloud Platform (GCP)

Projects

- Built a deep learning model with **U-Net** and **SimVP** for **22nd-frame prediction** and **segmentation** from the first 11 frames, achieving **0.455 IoU** using **pseudo-labeling** and **fine-tuning** to enhance accuracy on unlabeled data; Attained **1st position**. [Code] [Report]
- Formulated an SED system for DCASE Task 4A leveraging **bi-directional encoder representation from audio transformers (BEATS)** embeddings with **frequency dynamic convolutions (FDY-CNN)** and **asymmetric focal loss (AFL)**; Achieved **4th position**. [Report]