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 *(Yann LeCun)*, Natural Language Processing, Machine Learning, Computer Vision, Reinforcement Learning

Birla Institute of Technology and Science (BITS), Pilani **Aug 2017 - Jul 2021**
Bachelor of Engineering in Electrical Engineering ***GPA:*** *8.43/10.0 (First Division)* *Pilani, India*
Relevant Courses: Information Retrieval, Data Mining, Neural Networks and Fuzzy Logic, Object-Oriented Programming, Data Structures

Experience

Amazon **June 2024 - Aug 2024**
Applied Scientist Intern *Sunnyvale, California*

- Improved **real-time search and recommendation** for Amazon Music by leveraging a **transformer-based two-tower network** with **LoRA adaptors** and **mixture of experts (MoE)**, achieving a **61.5% accuracy boost** on low-performing queries to enhance relevance, user engagement, and satisfaction across **millions of searches**.
- Boosted **search ranking** performance with a **popularity-weighted contrastive loss** function, leading to a **25.17% increase in Recall@20** by incorporating **synthetic data generation** and **user search logs**.
- Deployed a **guardrailed** system on AWS, leveraging **large language model (LLM) fine-tuning** with **DPO** for user-aligned responses and **RAG** for key phrase generation and indexing from **lyrics**, achieving **sub-80ms** response time.

Music and Audio Research Lab (MARL) **Sep 2023 - Present**
Machine Learning Research Assistant *New York, USA*

- Developing a lightweight latent **diffusion transformer (DiT)** for **text-to-audio generation** on **resource-constrained devices**, leveraging **T5 embeddings**, **LoRA adaptors** and **post-training quantization (PTQ) methods** for efficient, high-quality synthesis.
- Contributed to open-source **Python** library: **Soundata** [Source] [Paper] with **47k 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, Singapore*

- 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%**.
- 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%**.
- Developed and **deployed** a scalable, **real-time** infant cry detection system [Paper] utilizing **depthwise-separable convolutions** and **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.

Publications

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

Technical Skills

Programming Languages: Python, C++, Java, JavaScript, Typescript, Matlab, Scala, ML
Tools: GitHub, Apache Kafka, Spark, Hadoop, Airflow, Kubernetes, Docker, Git, GPU, Nvidia CUDA, MySQL, JIRA
Frameworks: PyTorch, TensorFlow, PyTorch Lightning, Keras, OpenCV, Hugging Face Transformers, Scikit-learn, Matplotlib, NumPy, Pandas, ONNX, NLTK, spaCy, MLflow, XGBoost, Prometheus, Grafana, OpenAI API, Microsoft Azure, AWS (Sagemaker, Lambda, EC2)

Projects

- Built a deep learning model with **U-Net** and **SimVP** for **22nd- video frame prediction** and **segmentation** from first 11 frames, achieving **0.455 IoU** using **pseudo-labeling** and **fine-tuning** to enhance accuracy on unlabeled data; Attained **1st position**. [Code] [Report]
- Established a **multi-modal visual question answering (VQA)** system using **CNNs, LSTMs, and a stacked attention network (SAN)**, achieving state-of-the-art accuracy **80.5%** on the VQA-v2 dataset and the Radiology dataset. [Code] [Report]
- Formulated a sound event detection (SED) system leveraging **bi-directional encoder representation from audio transformers (BEATS)** embeddings with **frequency dynamic convolutions (FDY-CNN)**; achieved **4th position**. [Report]