Atharva Anand Joshi

atharva.joshi253@gmail.com • 412-954-7615 • https://www.linkedin.com/in/atharva-anand-joshi/ • https://atharva253.github.io/

RESEARCH INTERESTS

Speech Processing: Enhancement and Separation, Deep Learning, Representation Learning

EDUCATION

Carnegie Mellon University — Pittsburgh, PA
Master of Science in Electrical and Computer Engineering - Applied Advanced Study (AI/ML Systems) — December 2024
GPA: 4.0/4.0

Birla Institute of Technology and Science, Pilani — Pilani, India
Bachelor of Engineering in Electrical and Electronics Engineering — July 2022
GPA: 9.49/10
GRE: 331/340 (AWA: 4/6), TOEFL: 109/120

SKILLS

Programming: Python3, Bash Scripting, Java, C, CUDA , MATLAB
Deep Learning: PyTorch, TensorFlow, Keras, AWS and GCP for ML
Frameworks: ESPnet, Qualcomm DSP Toolkit, PySpark, WandB, OpenMP, LaTeX, Git

EXPERIENCE

HP Inc., Poly Advanced Technology Group — Austin, TX
Machine Learning Engineer 2 — February 2025 – Present
- Designing deep learning models for real-time use cases directly on Poly headset DSPs
- Prototyping and integrating ML-driven features into current and next-generation headset products

Machine Learning Intern — May – August 2024
- Introduced a novel ML pipeline that enhances speech through a combination of spatial and speaker representations
- Enabled adaptive voice personalization, allowing headsets to improve over time in isolating the user's speech
- Automated the complete development process to make it scalable to products with diverse form factors

Research and Development Intern — May – August 2023
- Developed deep learning models that can suppress stationary and impulsive background noises directly on headsets
- Optimized the inference pipeline on memory, compute and latency for real-time execution on the QCC5171 DSP
- Shipped the solution as the "Poly NoiseBlock AI for Headsets" feature in commercial products

American Express, Artificial Intelligence Labs — Gurgaon, India
Analyst - Product Development — July - December 2022
- Researched a blend of Tabular models with Tree-based algorithms for the Credit Default Prediction problem
- Enhanced model performance by leveraging extensively hand-engineered features, along with meta features
- Generated valuable business insights pertaining to features selection and effective aggregation of temporal features

Analyst Intern — January - June 2022
- Proposed a template-based journey that allows users to seamlessly create and deploy their machine learning pipelines
- Developed a framework that facilitates deployment of end-to-end Self Learning pipelines for Sequence Models

Adobe Research, India — Bangalore, India
Research Intern — May - August 2021
- Created rich user representations that can be projected onto edge servers, hence powering faster marketing services
- Performed various experiments around the extent of compression and updatability of the representations generated

PUBLICATIONS

- M. Someki, S. Bharadwaj, **A. A. Joshi**, C.J. Lin, J. Tian, J. Jung, M. Müller, N. Susanj, J. Liu, S. Watanabe. Context-Driven Dynamic Pruning for Large Speech Foundation Models. Proc. Interspeech 2025 (Accepted)

- **A. A. Joshi**, H. Settibhaktini, and A. Chintanpalli. Modeling concurrent vowel scores using the time delay neural network and multitask learning. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30:2452-2459, 2022
- **A. A. Joshi**, P. Bhardwaj, and S. M. Zafaruddin. Terahertz wireless transmissions with maximal ratio combining over fluctuating two-ray fading. IEEE Wireless Communications and Networking Conference, pages 1575-1580, 2022

PATENT

- S. Chakraborty, S. Choudhary, A. Sinha, S. Nair, M. Ghuhan, Y. Gagneja, **A. A. Joshi**, A. Tyagi, S. Gupta, "Generating Concise and Common User Representations for Edge Systems from Event Sequence Data Stored on Hub Systems", US12182829B2, Granted December 31, 2024

PROJECTS

Data Scalability Aspects for the Speech Enhancement Task
- Investigating large data (>10,000 hours) scalability aspects for the state-of-the-art speech enhancement models
- Compared several speaker-level data selection methods using diversity metrics based on speaker embeddings
- Extending the work by improving selection methods through non-intrusive speech quality prediction metrics

Query-Driven Dynamic Pruning for Large Speech Models
- Developed a novel frame-level gate prediction model which can dynamically prune speech LLMs
- Studied the impact of several context sources, including speaker characteristics and audio features on the Automatic Speech Recognition and Speech Translation capabilities of the model
- Analyzed the pruning pattern to gain insights on how the gate predictor decides which modules to prune

High Performance Parallel Implementations for Convolutional Neural Networks
- Provided fast OpenMP and CUDA implementations for various subroutines corresponding to the convolution layer
- Investigated several design choices in detail aiming to optimize speedup over a simple sequential C++ implementation
- Achieved maximum speedup of 4.23x on the Intel(R) Xeon(R) Silver 4208 CPU and 73.87x on the Nvidia Tesla T4 GPU

Proactive Servicing: American Express ML Challenge
- Utilized event sequences and demographic data to predict customer intent at the start of the Ask AmEx chat session
- Employed joint training of Bidirectional GRU with Feedforward Networks
- Attained a validation top-5 accuracy score of 0.768 and hence made it to the top 10 leaderboard out of ~80 teams

AWARDS AND RECOGNITION

HP InternStellar Awards: Technical Impact – 2nd Runner up among all US interns 2024
HP Intern Award – 4th position among ~90 engineering interns 2023
American Express Modeling Super Bowl – Top 10 Leaderboard 2022
OP Jindal Engineering and Management Scholarship 2020-2021
Institute Merit-Based Scholarship, Birla Institute of Technology and Science, Pilani – Top 2% 2019-2022

MENTORSHIP AND TEACHING

- **Graduate Teaching Assistant** for 10-605: ML with Large Datasets Spring 2024 - Fall 2024
  Instructors: Prof. Virginia Smith, Prof. Ameet Talwalkar and Prof. Geoff Gordon
- **Graduate Teaching Assistant** for 18-661: Introduction to ML for Engineers Fall 2023
  Instructors Prof. Yuejie Chi and Prof. Beidi Chen
- **Undergraduate Teaching Assistant** for BITS F312: Neural Networks and Fuzzy Logic Fall 2021
  Instructors: Prof. Surekha Bhanot and Prof. Bijoy Krishna Mukherjee

RELEVANT COURSEWORK

Graduate: Introduction to Deep Learning, Speech Recognition and Understanding, Machine Learning for Signal Processing, Deep Generative Modeling, Machine Learning with Large Datasets, Natural Language Processing, Introduction to Machine Learning for Engineers, CUDA Programming, Convex Optimization

Undergraduate: Neural Networks and Fuzzy Logic, Artificial Intelligence, Object Oriented Programming (Java), Digital Signal Processing, Digital Image Processing, Communication Systems, Operating Systems