Anand Mohan

Applied Scientist, Amazon AGI
London, UK

e-mail: moghan.anand@gmail.com
mobile : +44 77783 01937
linkedin.com/in/anandmoghan

Professional Summary

7+ years of applied research experience in large-scale Speech and Language AI. Currently leading the development of continual learning and hot-fixing pipelines for Amazon's frontier **speech-to-speech LLMs**, implementing **supervised fine-tuning**, **preference learning**, and **Constitutional AI-inspired prompt-based steering** to correct transcriptions and **align multi-billion parameter models** for Alexa+. Lead large-scale experimentation, custom evaluations, and model quality optimization while debugging complex training pipeline and model behavior issues. Expert in **parameter-efficient fine-tuning** and designing robust production systems. Proven ability to balance research innovation with operational reliability, translating cutting-edge alignment techniques into production while ensuring **responsible AI practices**.

Experience

- **Amazon** — Cambridge, UK
  Applied Scientist – AGI — *September 2024 – Present*
  - **Continual Learning and Hot-fixing (Science Lead)**: Leading the Continual Learning and Hot-fixing (Model Alignment) workstream within the Post-Training team for Nova Sonic, tailoring the model for Alexa+. Driving the complete science lifecycle – from architectural design and peer reviews to large-scale experimentation and evaluation, delivering critical **alignment and correction technologies for speech-to-speech LLMs** in production.
  - **Multi-Stage Post-Training Pipeline**: Key contributor to **post-training alignment of the Nova Sonic** foundation model for Alexa+ use cases. Working within complex multi-stage training pipelines involving **supervised fine-tuning** and **preference learning** to optimize model behavior for specific capabilities.
- **Amazon** — Hyderabad, India
  Applied Scientist – AGI (Alexa Speech) — *July 2019 – September 2024*
  - **Large Multilingual ASR and Continual Learning**: Key contributor to development and deployment of large multilingual ASR models in production. Designed and automated **parameter-efficient fine-tuning** pipelines using **LoRA adapters**, enabling efficient periodic model updates and customization while preserving core model integrity.
  - **High-Performance Data Pipelines**: Led design and development of critical data preparation infrastructure. Engineered highly optimized, **distributed pipelines handling millions of hours** of speech data, significantly reducing processing latency and eliminating technical debt in preparation for large-scale model training.
  - **Federated Learning Research**: Designed and built a **privacy-centric gRPC-based federated learning** system from scratch. Introduced novel convergence techniques for cross-silo federated training, demonstrating ability to implement complex training pipelines. Published findings at **ICASSP '21 and US Patent No. 12488798**.
  - **Language Modeling Innovation**: Developed and deployed multiple language modeling innovations, successfully migrating **production systems through technology generations** – from LSTM-based neural language models to advanced BERT-based architectures – while maintaining production stability and performance.
  - **Research-to-Production Engineering**: Owned **end-to-end production responsibilities** across multiple ASR generations (Hybrid ASR → RNN-T → Multilingual ASR), successfully **translating research techniques into production-ready systems**. Drove engineering excellence through **process automations**, technical debt elimination, and reliable deployment practices. Debugged **complex training pipelines and model behaviors** under **time-critical production** conditions.
  - **Cross-Functional Collaboration and Mentoring**: Collaborated extensively **across research and engineering teams** to translate research innovations into production. Mentored junior scientists and **led 80+ technical interviews**, contributing to team growth and maintaining high hiring standards.
- **Zoho** — Chennai, India
  Member of Technical Staff – Site24×7 — *July 2015 – June 2016*
  - **Full-stack Development**: Designed and optimized robust Java-based server APIs and built AngularJS web clients. Led end-to-end deployment of 8 major production releases for real-time performance monitoring service, demonstrating **early experience with production systems at scale**.

Skills

- **Specializations**: Post-Training (Supervised Fine-Tuning, Preference Learning), Parameter-Efficient Fine-Tuning (LoRA), Large Language Models, Distributed Training, Model Alignment, Production ML Systems, ASR, SID, LID
- **Frameworks**: PyTorch, PySpark, TensorFlow, DeepSpeed, Megatron-LM, Kaldi-ASR

Publications

Google Scholar: scholar.google.com/citations?user=D1KIt3cAAAAJ

- **AMuSE: Attentive Multilingual Speech Encoding for Zero-Prior ASR**
  Accepted at IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) *2025*
- **Cross-silo Federated Training in the Cloud with Diversity Scaling and SSL**
  IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) *2021*
- **Towards Relevance and Sequence Modeling in Language Recognition**
  IEEE Transactions on Audio, Speech and Language Processing *2020*
- **Attention based Hybrid i-vector BLSTM Model for Language Recognition**
  Annual Conference of the International Speech Communication Association (INTERSPEECH) *2019*
- **End-to-End Language Recognition Using Attention Based Hierarchical GRU Models**
  IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) *2019*
- **The LEAP Speaker Recognition System for NIST SRE 2018 Challenge**
  IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) *2019*

Patents

- **Continuous Learning for Machine Learning Models**
  US Patent No. 12488798 *2025*

Education

- **Indian Institute of Science (IISc), Bangalore** Bangalore, India
  Master of Technology (M. Tech) in Artificial Intelligence *2017 - 2019*
- **National Institute of Technology (NIT), Calicut** Calicut, India
  Bachelor of Technology (B. Tech) in Electronics and Communication Engineering *2011 - 2015*

Projects

- **Attention Based Relevance Modelling for Speaker and Language Recognition** Master's Thesis
  Guide: Dr. Sriram Ganapathy, Indian Institute of Science, Bangalore *May 2018 – June 2019*
  - **Attention Mechanisms for Sequence Modeling**: Designed and analyzed frame-level and segment-level attention mechanisms to identify relevant information in variable-length sequences. Demonstrated through **ablation studies and weight visualization** that attention significantly outperforms uniform pooling, particularly in noisy conditions – a core principle **fundamental to modern LLMs**.
  - **Language Recognition with Hierarchical Modeling**: Developed i-BLSTM and hierarchical GRU architectures with attention for dialect identification. Achieved 13-17% relative improvement over baselines on NIST LRE 2017 in noisy and multi-speaker conditions by dynamically weighting relevant speech segments.
  - **Speaker Verification with Multi-Head Attention**: Implemented multi-head attention for speaker embeddings, achieving 38% relative improvement on multi-speaker VAST corpus. Analysis showed attention learns to suppress silence and irrelevant regions without explicit supervision.
- **Sparse Representation & Recovery of Graph Signals** Bachelor's Thesis
  Guide: Dr. G. Abhilash, National Institute of Technology, Calicut *Aug. 2014 – May 2015*
  - **Spectral Graph Theory and Signal Processing**: Studied spectral graph theory and implemented 2-channel perfect reconstruction wavelet filter banks using bipartite graphs for multi-resolution analysis.

Leaderships and Achievements

- **Reviewer: ICASSP, SPCOM, Amazon Machine Learning Conference (AMLC)**
  Peer reviewer in international conferences in Speech and ML *2020 - 2025*
- **Primary Contributor: Coswara – Cough and speech based Covid '19 detection tool**
  Indian Institute of Science, Bangalore. Website: coswara.iisc.ac.in *2020 - 2021*
- **Student Placement Coordinator**
  Indian Institute of Science, Bangalore *2018 - 2019*
- **Branch Representative (Elected)**
  Students Affairs Council, National Institute of Technology, Calicut *2014 – 2015*
- **Manager, Registration Committee**
  Tathva & Ragam, techno-management & cultural festivals, National Institute of Technology, Calicut *2013 - 2014*