Akash Mahajan

+1-(650)-546-5305 akashmjn.1@gmail.com akashmjn.me LinkedIn/akashmjn

SUMMARY

I currently work as an applied scientist with expertise in speech recognition, digital signal processing, and deep learning. I most recently worked on transcription & diarization for AI-assisted meeting recap.

I do my best work in environments that require quick learning from experts, simplifying the complex, and a product-minded approach.

EXPERIENCE

Senior Applied Scientist | **Microsoft Azure Speech** Nov 2018-present

Offline speech recognition

- Developed and shipped a Conformer-S2S based system to Azure Batch API that improved accuracy by 10% at 15% lower cost (real-time factor)
- Worked across stack: data pipeline, distributed training, optimizing CPU/GPU inference in ONNX/C++

Diarized, farfield transcription of conference rooms

- Co-developed metrics, led system-wide error analysis locating bottlenecks in complex multi-component pipeline (mic array processing, speech separation, hybrid ASR, diarization, display formatting)

Real-time speech recognition

- Developed and shipped LC-BLSTM based hybrid (HMM / WFST) acoustic models, incorporating larger offline teacher models into production recipes

Founding Data Science member | **Ather Energy** [pioneering Indian EV startup] 2015-2016

- Co-developed a roadmap for vehicle intelligence by prototyping features (e.g. speed bump detection, predictive maintenance)
- Conceptualized and led the engagement for a unique EV test ride visualization experience [blog]

PROJECTS

tinydiarize: Minimal extension of Whisper for diarization [github] 2023

- Built an extension of OpenAI's Whisper ASR model for speaker diarization with special tokens
- Released with integration in whisper.cpp runnable on MacBooks/iPhones

OrcaHello: AI-assisted 24x7 hydrophone monitoring | **Microsoft Global Hackathon** 2019-2022

- Co-founder and lead for award-winning project ($30k+ Azure credits, non-profit cash grants)
- Went from no labeled data to live inference on Azure at 3 hydrophone locations, with feedback from moderator web-app. Co-designed and built with team of 20+ volunteer hackers [github]

Attention I'm trying to speak: Text to speech synthesis | **Stanford [github]** 2018

- Built a low-cost convolution-attention based system trainable in half a day on a K80 GPU
- Awarded best poster amongst 50+ projects in CS224n (Deep Learning for NLP)

EDUCATION

Stanford University, M.S. Management Science & Engineering 2016-2018

Deep Learning/Digital Signal Processing, Databases/Computer Systems, Marketing/Strategy/Design
CA (course assistant) for Machine Learning (CS229) & Deep Learning (CS230)

Indian Institute of Technology, Madras, B.Tech. 2011-2015

Chemical Engineering, minor: Control Systems (linear algebra, stats, signal processing)

MISC

Patents

- US11044287B1 Microsoft (granted, 2021)
  Caption assisted calling to maintain connection in challenging network conditions
- WO2018020475A1 Ather Energy (PCT application, 2018)
  A method and system for determining an operational condition of a vehicle component

Programming languages: Python, C/C++, HTML/CSS/JS/C# (hackable)
Frameworks & Tools: Pytorch distributed, ONNX runtime, Docker, AzureML/Batch/Blob data pipelines