

MOHAMMAD OMAR KHURSHEED

Applied Scientist – AI Safety & Speech

omarkhu@gmail.com — 413-230-8914 — linkedin.com/in/mkhursheed — github.com/omarkhursheed — omarkhursheed.github.io

RESEARCH INTERESTS

Machine learning for speech and multimodal assistants; low-latency and on-device models; evaluation and robustness of audio and LLM systems; fairness and bias in speech/ASR; mechanistic interpretability for understanding model behavior, computational social science for understanding bias in large-scale datasets

SELECTED PUBLICATIONS

- Two Centuries of Women’s Voices: Computational Analysis of Gender Participation and Suffrage Debates in British Parliament **Khursheed MO** (co-first author), Sawkar M., Khudabukhsh AR (*in submission at WWW*), 2025
- Small Footprint Slimmable Networks for Keyword Spotting. Akhtar Z, **Khursheed MO**, Du D, Liu Y. *ICASSP*, 2023.
- Latency Control for Keyword Spotting. Jose C, Wang J, Strimel GP, **Khursheed MO**, Mishchenko Y, Kulis B. *Interspeech*, 2022.
- Tiny-CRNN: Streaming Wakeword Detection in a Low Footprint Setting. **Khursheed MO**, Jose C, Kumar R, Fu G, Kulis B, Cheekatmalla SK. *ASRU*, 2021.
- Analyzing Gender Bias within Narrative Tropes. Gala D, **Khursheed MO** (co-first author), Lerner H, O’Connor B, Iyyer M. *NLP+CSS Workshop @ EMNLP*, 2020.

EMPLOYMENT AND RESEARCH EXPERIENCE

Amazon – Alexa Edge AI

Applied Scientist II

Sunnyvale, CA

December 2022 – Present

- Developed training strategies for underperforming customer cohorts and trained wakeword models across multiple regions, achieving a 10% relative FRR reduction for targeted cohorts on Echo devices.
- Led a project to predict device-directedness using late fusion of audio metadata (e.g., ERLE) and neural network posteriors to reduce self-wakes by 50% during natural conversation.
- Designed a dynamically slimmable transformer architecture for wakeword detection on low-power devices, enabling parallel training of models with different compute/accuracy tradeoffs.
- Investigated cross-architecture knowledge distillation (transformer to CNN) for multi-wakeword detection to improve performance in constrained speech settings.

Applied Scientist I

January 2021 – December 2022

- Led a small team of scientists to develop low-power wakeword models for mobile devices and earbuds, improving previous models by up to 60% relative FRR across multiple locales using semi-supervised learning on large unlabeled datasets.
- Shipped production models for Echo Buds and mobile apps with calibrated thresholds and monitoring; received the Amazon Deliver Results Award (Q1-Q2 2021) for a critical product launch to 10M+ users.

Applied Scientist Intern

May 2020 – September 2020

- Developed small-footprint convolutional recurrent neural networks (CRNNs) for keyword spotting, improving accuracy over CNNs by about 25% with roughly 10% fewer parameters; prototyped on-device inference via parallel LSTMs (work published at ASRU 2021).

RIT (Prof. Ashiqur Rahman KhudaBukhsh)

Research Collaborator

Remote

January 2024 – Present

- Built an analysis pipeline over 200+ years of UK Hansard debates to study gender representation and discourse patterns in parliamentary speech.
- Released tools for exploring and visualizing the corpus (hansard-nlp-explorer); ongoing work on connecting historical trends to modern language model behavior.

NGRAM Lab (Prof. Mohit Iyyer), UMass Amherst

Graduate Student Researcher

Amherst, MA

January 2020 – December 2020

- Curated a large-scale dataset linking narrative tropes with IMDb and Goodreads metadata to analyze gender bias in narrative fiction.
- Co-developed a “genderedness” score and studied its correlation with ratings and creator gender, leading to a publication at the EMNLP NLP+CSS workshop.

- Implemented components of a face-recognition and landmark feature extraction system that mapped audio/visual signals to fNIRS brain activity to study emotion regulation.

AI SAFETY AND ALIGNMENT EXPERIENCE

- **Bluedot Impact: AI Safety Fundamentals** (July–October 2024). Final project on mechanistic interpretability for Whisper-style models using sparse autoencoders to probe accent-related features and study potential bias in speech recognition.
- **ARENA 3.0**. Completed the public curriculum via self-study, covering transformer internals, sparse autoencoders, interpretability techniques, and alignment methods.

EDUCATION

University of Massachusetts Amherst
M.S. in Computer Science

2019 – 2021
GPA: 3.97/4.0

Aligarh Muslim University, Aligarh
B.Tech. in Computer Engineering

2015 – 2019
GPA: 9.52/10.0

Selected achievements: Sir Syed Global Scholar Award; ICPC Regionals; AMU-OSS Co-founder; Hack36 2018 Top 12.

SERVICE

- Program committee / reviewer: NeurIPS 2024; ICML 2025; AMLC 2023–2025; SIGIR 2022; JMIR, AAAI 2026