

Ramin Anushiravani

New York City, NY | [Linkedin](#) | [Github](#) | [Website](#) | ramin.audio@gmail.com

Skills

- **Deep Learning Frameworks:** PyTorch, TensorFlow, Keras, TFLite, Sklearn, HuggingFace
- **Foundation and Multimodal AI:** Classical ML, BERT, GPT, Reinforcement learning, Llama, AudioLM, Vision Transformer, Swin, ViViT, EfficientNet, , Flamingo, Whisper, Audio Spectrogram Transformer, Audio tokenizer, Wav2Vec, Conformer, YamNet
- **Search:** RAG, Vectorized Search, Entity recognition, query understanding, recommendation systems
- **Model optimization:** Self-supervised and contrastive learning, LoRA, Few-shot, prompt engineering and instruction fine-tuning, prompt engineering, quantization, knowledge distillation, pruning
- **Audio Processing:** Signal processing, Blind source separation (NMF), dereverberation, denoising, feature engineering, 3D audio
- **MLOps and deployment:** AWS (S3, EC2, SageMaker Pipelines), MLFlow, Flask, FastAPI, GitHub Actions, Dockers, Optuna

Experience

Precision Neuroscience, New York, NY – *Staff Machine Learning Scientist* – 11/2023 to Present

- Implemented a novel **transformer-based multitask foundation model** from ECoG data pretrained using a **self-supervised contrastive objective** and fine-tuned on supervised tasks to produce high quality embeddings.
- Built scalable and reusable **machine learning and signal processing pipelines** to process terabytes of high-dimensional time series data.
- Fine-tuned SOTA **ASR models** to annotate speech data collected from operating rooms and align it with neural data.
- Developed **model interpretability tools** using saliency and attention maps to assess electrode contributions to decoding.
- **Optimized model latency** on NVIDIA Orion Nano by 24x while maintaining performance leveraging neural architecture search and quantization.
- Developed real-time **few-shot inference** models for hand gesture classification from motor cortex activity, achieving 85% F1 score and regression for real-time cursor control achieving 79% R^2 in the operating room.

United HealthGroup, San Mateo, CA – *Sr Principal ML Engineer* – 01/2021 to 10/2023

- Led a team of data and ML engineers to develop, launch, and maintain **text understanding models for consumer search** products.
- Developed and maintained multilingual auto-correct using character-level bidirectional LSTMs and N-grams.
- Developed auto-complete and auto-suggest algorithms using FSTs and **fine-tuned GPT-2 on healthcare queries**.
- Created AI-powered search tools **serving 40 million active members** directly driving significant improvements in click-through rates and user satisfaction, leading to a 5x increase through A/B testing.
- Pre-trained and fine-tuned several encoders (BERT, RoBERTa, DistillBERT) to generate **sentence embeddings** to enable vectorized search functionality and other downstream tasks such as entity recognition.
- Benchmarked **ASR models**, including wav2vec 2.0 and NVIDIA NeMo, and deployed **conversational AI agents** for call routings and abstractive summarization using T5, enhancing customer service efficiency.

CurieAI, Menlo Park, CA – Machine Learning Scientist – 04/2018 to 01/2021

- Developed novel hybrid on-device and cloud **audio understanding** for monitoring chronic respiratory diseases in challenging acoustic environments, achieving an 80% increase in recall and an 86% improvement in precision over existing licensed models.
- Spearheaded **machine learning life cycles**, from data collection and annotation to signal processing and continuous model training, driving significant improvements in model performance and efficiency.
- Developed an AI-driven course of action recommendation system, leveraging patient history and engagement data.

DSP Concepts, Santa Clara, CA – Algorithm Engineer – 09/2017 to 04/2018

- Engineered **noise reduction and dereverberation algorithms** for improving wake-word detection on smart speakers.
- Automated testing protocols for audio algorithms, ensuring robust performance across various acoustic conditions.

Dolby Labs, San Francisco, CA – Audio Engineer – 09/2016 to 09/2017

- Developed an automated system for detecting infringements of Dolby **audio codecs**.
- Delivered expert tutorials and white papers on cutting-edge audio processing and deep learning, educating senior executives on emerging technologies.
- Managed extensive patent portfolio, drafting claims and responding to complex office actions.

Prior roles: [Adobe \(Audio editing\)](#), [GN-ReSound \(Hearing aids\)](#), [Advanced Digital Science Center \(Microphone arrays, Singapore\)](#)

Written Work & Publications

- Granted: [Sound Enhancement through Reverberation Matching](#)
- Granted: [Methods for Explainability of Deep-Learning Models](#)
- Granted: [Intelligent Health Monitoring](#)
- Granted: [Design of Stimuli for Symptom Detection](#)
- Pending: [Domain aware autocomplete](#)
- Pending: [Graph-based data compliance using natural language text](#)
- Pending: [Interactive map-based visualization system related to multichannel search for complex search domains](#)
- Pending: [Machine learning techniques for generating domain-aware query expansions](#)
- Pending: [Multi-channel search and aggregated scoring techniques for complex search domains](#)
- Pending: [Text embedding-based search taxonomy generation and intelligent refinement](#)

What is attention?	How does ChatGPT work?	Bard - Google's Response to ChatGPT
3D Audio	3D Audio for single-channel audio using visual cues	Sound Source Localization
Model Optimization	AI summaries	Seamless Acoustics Matching of Disparate Recordings
Example Based Audio Editing	A computer vision approach to speech enhancement	

Education

08/2011 - 12/2016

M.S. & B.S., Electrical & Computer Engineering, University of Illinois at Urbana-Champaign (GPA: 3.97/4, 3.86/4)