Ramin Anushiravani

New York City, NY | Linkedin | Github | Website | ramin.audio@gmail.com

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)*

Skills

- **Deep Learning Frameworks:** PyTorch, TensorFlow, Keras, TFLite, Sklearn
- **Foundation and Multimodal AI:** Classical ML, BERT, GPT, Reinforcement learning, Llama, AudioLM, Vision Transformer, EfficientNet, Wav2Vec, Conformer
- **Search:** RAG, Vectorized Search, Entity recognition, query understanding, recommendation systems
- **Model optimization**: Self-supercised 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

Experience

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

- Built scalable and reusable machine learning and signal processing pipelines to process terabytes of brain signal (ECoG) data, facilitating processing, visualization, model training, and hyperparameter tuning.
- Designed hand-crafted features engineered to capture meaningful spatiotemporal neurophysiological signals.
- Implemented a novel multitask foundation model pretrained on self-supervised contrastive pretraining and fine-tuned on supervised high-dimensional time series neural recordings.
- Deployed error corrections techniques utilizing language models to decode speech from brain activity.
- Developed model interpretability tools using saliency and attention maps to analyze the contribution of electrodes on decoding.
- Deployed neural foundation model in real-time enabling few-shot inference using neural embeddings in operating rooms.
- Conducted ablation studies and hyperparameter tuning using Optuna, optimizing model architectures and training configurations.
- Developed real-time hand gesture classification models from motor cortex activity, achieving 85% F1 score in operating rooms.
- Developed regression models to decode motor cortex activity for real-time cursor control achieving 79% R^2 in OR.
- Collaborated cross-functionally with neuroscientist, product, and ML engineering teams to develop comprehensive brain-computer interface solutions.

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 a character level bidirectional LSTMs and N-grams and auto-complete and auto-suggest algorithms using FSTs and GPT-2 for multiple customers, serving 40 million active members and achieving significant improvements in click-through rates and user satisfaction resulting in a 5x increase in user engagement validated through A/B testing.

- Fine-tuned several encoders (BERT, RoBERTa, DistillBERT) and decoders (GPT2, Llama) LLMs on healthcare data on downstream tasks such as entity and intent recognition.
- Developed healthcare sentence embeddings to enable vectorized search for a consumer facing website.
- Benchmarked ASR models wav2vec2.0 and NVIDIA NeMo and deployed conversational AI agents for directing calls and abstractive summarization using T5, and improving customer service efficiency.

Curie AI, Menlo Park, CA – *Machine Learning Scientist* – 04/2018 to 01/2021

- Developed novel audio understanding models 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
- Contributed to investor pitch decks and drafting patents, assisting in securing funding and protecting intellectual property.

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 (algorithms microphone arrays, Singapore)

Written Work & Publications

Patents

- 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

Papers/Slides

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

Learning

- Udacity: Generative AI Nanodegree (2025), NLP Nanodegree (2021), ML DevOps Engineer (2022), Deep Reinforcement learning (2023)