Arsh Zahed

Machine Learning Engineer

Objective: $\max \mathbb{E}\left[\|\text{Experience}\|^2 + \|\text{Knowledge}\|^2\right]$
Optimization Method: Full-Time Engineer/Researcher

azahed98.github.io | azahed98 | arsh-zahed | ********** | **********

Experience

NVIDIA | Deep Learning Engineer
AI Applications | July '20 – Current

- Deployed model conversion tool for Riva Speech Services to optimize models for server deployment with Triton using ONNX and TensorRT. Supports over 15 different pipelines, accelerating for >12x.
- Designed and built TAO-LM, tool for training/tuning N-Gram models.
- Standardized testing framework to increase coverage from 40% to 72%.
- Contributed to online demo for Riva, used by >500 users a day.

BERKELEY AI RESEARCH | Researcher & Grader
AutoLab | Jan '19 - Jan '20

- Research in Reinforcement, Imitation and Online Learning.
- Reduced failure of safety using uncertainty estimation by 14%.
- Built 8 experiments for imitation learning with an improving supervisor.
- Graded/tutored for Deep Learning and Optimization for >600 students.

GOOGLE | Software Engineer Intern
Chrome Media Audio | May '18- Aug '18

- Created TF Estimators experimentation framework to predict the speech coding quality of WaveNet/Lyra while reducing bitrate by 50%.
- Collected 7000 user-rated WaveNet samples. Ran experiments with RNNs, Dilated Convolutions and Variational Autoencoders.

LAUNCHPAD | President & Project Lead
UC Berkeley Student Org | Jan '17 - May '20

- As president, led and organized educational ML workshops and meetings for 40+ members. Maintained relationships with 3 sponsors.
- As PL, led 16 developers on 2 research-oriented projects.

Publications

"On-Policy Imitation Learning from an Improving Supervisor"

- Conference on Robot Learning (CORL), 2019
- Real World Sequential Decision Making Workshop at ICML, 2019.

Projects

UNCERTAINTY AWARE PHYSICS ESTIMATION Python, PyTorch | 2021

- Used uncertainty estimation to create an active learning framework for physics estimation. Achieved a >50% decrease in required data.

EXPRESSIVE TTS FROM INFERRED EMBEDDINGS Python, PyTorch | 2020

- Inferred style-embeddings from text to improve generated speech.
- Improved F0 Frame Error by 8% with audible improvement.

METAL - MAML EXPLORATION WITH METRICS Python, TensorFlow | 2019

- Developed Policy Metrics to form a topology on the set of policies of an MDP that helps guide task-specific exploration.
- Used with imitation learning for 22% reduction in training speed.

SEMI-GENERALIZED GRASPING Python, ROS | 2018

- Built an algorithm to sense and grasp using a Kinect and Sawyer robot.

Skills

TOPICS & FIELDS
Deep Learning • Speech Processing • Natural Language Processing • Digital Signal Processing • Generative Models • Reinforcement Learning

PROGRAMMING
Python • C • C ++ • JavaScript • R • Java • Protobuf • Bash • LaTeX

LIBRARIES & TOOLS
TensorFlow • PyTorch • Triton • AWS • GCP • Docker • Kubernetes

Education

STANFORD UNIVERSITY
Non-Degree | Sep '21 - Present
CS | GPA: 3.7

UC BERKELEY
B.S. | Aug '16 - May '20
EECS | GPA: 3.703

Coursework

STANFORD

CS 224n	Natural Language Processing
CS 236	Deep Generative Models

UC BERKELEY

CS 194-26	Quantum Computing
CS 191	Computational Photography
CS 189	Machine Learning
CS 188	Artificial Intelligence
CS 170	Algorithms
CS 162	Operating Systems
CS 161	Computer Security
EE 225b	Digital Image Processing
EE 127	Convex Optimization
EE 126	Probability Theory
EE 123	Discrete Signal Processing
EE 120	Signals & Systems
EE 106a	Robotics
Math 141	Differential Topology
Math 110	Linear Algebra
Math 104	Real Analysis
Music 108	Music Cognition
Stat 154	Stochastic Processes
Stat 153	Time Series Analysis