Vashisht Madhavan

Machine Learning and Artificial Intelligence

vashisht.madhavan@gmail.com vashishtmadhavan.github.io https://www.linkedin.com/in/vashishtmadhavan

Experience

Humanlike (YC S23)

Founder, CTO — *May 2023 - Present*

- Leveraged LLMs and generative speech models to build low-latency, realistic AI voice bots that automate phone calls for businesses.
- Grew the company from 0 to $200k ACV, landing large enterprise customers such as Opendoor.
- Led engineering and product from the ground up, deploying streaming and pipelining architectures on AWS to handle 1000s of calls a month.

Snorkel AI

Staff Machine Learning Engineer — *Jun 2022 - May 2023*

- Led development for the structured document NLP product, adding OCR, feature extraction, and data labeling features to build custom ML models.
- Built platform integrations with Snowflake, Azure, and Databricks, leading to official partnerships with each company.
- Integrated foundation model workflows like embedding search, prompt-based labeling, and language model fine-tuning into the platform, leveraging document-specific models like LayoutLM.

Senior Machine Learning Engineer — *Jun 2021 - Jun 2022*

- Led teams of 3+ engineers to build high-value productivity features like AutoML, labeling suggestions, and document search.
- Helped 2 clients build custom document NLP models during 4-week POCs, which converted to contracts worth $1M+ annually.

Element Inc.

Research Engineer — *Nov 2019 - Jun 2021*

- Reduced memory and latency of facial recognition models by 50% by leveraging novel network compression techniques.
- Built de-duplication services for large facial image databases using embedding-based clustering approaches.
- Developed document annotation tools to train and deploy on-device OCR models for passport/ID verification.

Uber Technologies Inc.

Research Scientist II — *Jun 2017 - Nov 2019*

- Published 4 research papers on deep reinforcement learning with a focus on exploration, generalization, and fast adaptation.
- Reduced latency by 12% for autonomous vehicle object detection models using neural architecture search.
- Applied reinforcement learning to Uber Freight surge pricing, leading to 15% reduction in shipping costs.

Berkeley AI Research Lab

Graduate Student Researcher — *Jan 2016 - May 2017*

- Released the BDD100k dataset, a large collection of annotated driving videos, and annotation tools for labeling driving data.
- Leveraged semi-supervised transfer learning to adapt vehicle perception models from synthetic urban settings to real-world scenarios.

Research

DECORE: Deep Compression with Reinforcement Learning — *CVPR 2022*

- Manoj Alwani, Yang Wang, Vashisht Madhavan

BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling — *CVPR 2020*

- Fisher Yu, Wenqi Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, Trevor Darrell

Improving Exploration in ES for Deep RL via a Population of Novelty-Seeking Agents — *NeurIPS 2018*

- Edoardo Conti*, Vashisht Madhavan*, Felipe Petroski Such, Joel Lehman, Kenneth O. Stanley, Jeff Clune

Deep Neuroevolution: GAs are a Competitive Alternative for Training Deep NNs for RL — *NeurIPS 2018 Workshop*

- Felipe Petroski Such, Vashisht Madhavan, Edoardo Conti, Joel Lehman, Kenneth O. Stanley, Jeff Clune

Best Practices for Fine-Tuning Visual Classifiers to New Domains — *ECCV 2016 Workshop*

- Brian Chu*, Vashisht Madhavan*, Oscar Beijbom, Judy Hoffman, Trevor Darrell

Education

University of California, Berkeley — *Aug 2016 - May 2017*

M.S. Electrical Engineering and Computer Science .

Computer Vision and Machine Learning - Advisor: Trevor Darrell

University of California, Berkeley — *Aug 2012 - May 2016*

B.E. Electrical Engineering and Computer Science .

Skills

Languages & Technologies

- Python, Javascript, C++, PostgreSQL AWS, Snowflake, Git, NextJS, FastAPI

Machine Learning

- PyTorch, HuggingFace, Spark, Scikit-Learn, SpaCy, Azure AI SDKs