

VASHISHT MADHAVAN

✉ vashisht.madhavan@gmail.com

🏠 vashishtm.com

🌐 <https://www.linkedin.com/in/vashishtmadhavan>

Experience

Pika

RESEARCH ENGINEER

Jan 2024 - Present

- Leading data curation efforts for large-scale pretraining of video generation models.
- Responsible for post-training experimentation and infrastructure, allowing the team to ship 10+ user-facing fine-tunes (i.e. Pikaffects).
- Fine-tuned multi-modal language models for video captioning, categorization, and retrieval, allowing the team to build higher quality datasets.

Humanlike (YC S23)

FOUNDER, CTO

May 2023 - Dec 2023

- Leveraged LLMs and generative speech models to build low-latency, realistic AI voice bots that automate phone calls for businesses.
- Grew revenue from 0 to \$200K annually, landing large enterprise customers like Opendoor.
- Led engineering and product from the ground up, scaling systems to handle 1000s of phone calls per month.

Snorkel AI

STAFF MACHINE LEARNING ENGINEER

Jun 2022 - May 2023

- Tech lead for the structured documents product, used by 5+ F500 customers to extract, label, and classify data from PDFs.
- 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 a team of 3+ engineers to build productivity features like AutoML, labeling suggestions, and document search.
- Helped close \$1M+ in ACV by assisting 2 clients in delivering custom document NLP models during 4-week POCs.

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 that focus on exploration, generalization, and fast adaptation.
- Reduced inference latency by 12% for autonomous vehicle object detection models using neural architecture search.
- Applied reinforcement learning to Uber Freight surge pricing, leading to a 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.

Research

AuroraCap: Efficient, Performant Video Detailed Captioning and a New Benchmark

ICLR 2025

- Wenhao Chai*, Enxin Song*, Yilun Du, Chenlin Meng, Vashisht Madhavan, Omer Bar-Tal, J.N. Hwang, Saining Xie, Chris D. Manning

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

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, Kubernetes, AWS, BigQuery, Snowflake, Git, NextJS, FastAPI

Machine Learning

- PyTorch, NumPy, HuggingFace, Ray, Scikit-Learn, SpaCy