

Yash Savani

Ph.D. Computer Science, Carnegie Mellon University

<https://yashsavani.com>
<https://github.com/yashsavani>
ysavani@cs.cmu.edu
@yashsavani_

Professional Statement

I am a research scientist and engineer with expertise in applied mathematics, algorithms, statistics, and modern machine learning. My work focuses on statistical learning, optimization, differentiable programming, and cutting-edge technologies like diffusion models, large language models (LLMs), and CUDA programming. I aim to develop efficient, scalable, and impactful solutions to complex computational problems.

Education

Carnegie Mellon University, Pittsburgh, PA
Ph.D. Computer Science (GPA 0.00 / 0.00)

Aug 2021 - Current

Stanford University, Palo Alto, CA
M.S. Statistics (GPA 3.77 / 4.00)
B.S. Computer Science (GPA 3.54 / 4.00)

Mar 2015 - Jun 2017
Sep 2013 - Jun 2017

Conference Publications

Diffusing Differentiable Representations.

Y. Savani, M. Finzi, J. Z. Kolter

Advances in Neural Information Processing Systems (NeurIPS) 2024.

Deep Equilibrium Optical Flow Estimation.

S. Bai, Z. Geng, Y. Savani, J. Z. Kolter

IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR) 2022.

NAS-Bench-x11 and the Power of Learning Curves.

S. Yan, C. White, Y. Savani, F. Hutter.

Advances in Neural Information Processing Systems (NeurIPS) 2021.

Exploring the Loss Landscape in Neural Architecture Search.

C. White, S. Nolen, Y. Savani.

Conference on Uncertainty in Artificial Intelligence (UAI). PMLR, 2021.

BANANAS: Bayesian Optimization with Neural Architectures for Neural Architecture Search.

C. White, W. Neiswanger, Y. Savani.

Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) 2021.

Intra-Processing Methods for Debiasing Neural Networks.

Y. Savani, C. White, and N. S. Govindarajulu.

Advances in Neural Information Processing Systems (NeurIPS) 2020.

A Study on Encodings for Neural Architecture Search.

C. White, W. Neiswanger, S. Nolen, Y. Savani.

Selected for Spotlight Presentation

Advances in Neural Information Processing Systems (NeurIPS) 2020.

Workshop Publications

A Study on Encodings for Neural Architecture Search.

C. White, W. Neiswanger, S. Nolen, Y. Savani.
ICML Workshop on AutoML, 2020.

Local Search is State of the Art for Neural Architecture Search Benchmarks.

C. White, S. Nolen, Y. Savani.
ICML Workshop on AutoML, 2020.

Neural Architecture Search via Bayesian Optimization with a Neural Network Prior.

C. White, W. Neiswanger, Y. Savani.
NeurIPS Workshop on Meta Learning, 2019.

Deep Uncertainty Estimation for Model-based Neural Architecture Search.

C. White, W. Neiswanger, Y. Savani.
NeurIPS Workshop on Bayesian Deep Learning, 2019.

Teaching

TA Fall 2023	15-750: Algorithms in the Real World, Fall 2023 (CMU) Instructor: Rashmi Vinayak
NeurIPS 2020 Short Presentation Oct 2020	Intra-Processing Methods for Debiasing Neural Networks. https://www.youtube.com/watch?v=PcCj91K7jO0
Abacus.AI Workshop Aug 2020	Explainability and bias in Neural Nets. https://www.youtube.com/watch?v=0PO6DgQe_9M
AICamp Workshop May 2020	Unsupervised Learning & Deep Learning Based Forecasting. https://www.youtube.com/watch?v=RcoW1KxXezE
Abacus.AI Talk Feb 2020	Anomaly Detection. https://www.youtube.com/watch?v=jDjwY9bB7Ec
Abacus.AI Talk Aug 2019	XLNET: The State-of-the-Art in Language Models. https://www.youtube.com/watch?v=jh81xHY6uBw

Experience

Research Scientist May 2020 - May 2021	Abacus.AI , San Francisco, CA I performed research in the AutoML / NAS and Fairness in ML domains. We wrote five papers based on this work.
Machine Learning Engineer Apr 2019 - May 2020	Abacus.AI , San Francisco, CA I designed and implemented scalable deep learning architectures including LSTM forecasting models, AutoML / NAS regression and classification models, GAN data augmentation models, and VAE anomaly detection models among others.
Machine Learning Engineer Aug 2017 - Dec 2018	Primer Technologies, Inc. , San Francisco, CA I worked on improving contemporary statistical learning and applied graph theory models for natural language applications. The machine intelligence algorithms I developed help decipher global news data.
Research Intern Jul 2015 - Sep 2015	Andrew Ng's Artificial Intelligence Lab , Stanford University, CA I worked on the system infrastructure and CUDA code for a hybrid CNN and LSTM architecture designed to instantly detect and semantically segment images and videos with multiple stimuli.

Cofounder (CTO) Jul 2014 - Dec 2015	Ebotic , Palo Alto, CA I worked with an international team to develop an intelligent drone platform that applied advanced flight technologies, SLAM, and deep learning for improved flight stability and awareness.
Research Intern Jun 2014 - Aug 2014	Sebastian Thrun's Artificial Intelligence Lab , Stanford University, CA I improved the performance of machine learning algorithms for smart home applications by adding thermal image descriptors into a robotics pipeline.

Skills

Computer Languages	Python , Julia, C / C++, CUDA, Javascript, R, Java, MATLAB, Haskell, LaTeX, SQL, NoSQL, and HTML5 / CSS3.
Frameworks / Tools	PyTorch , JAX, TensorFlow, HuggingFace , NumPy , Matplotlib , Jupyter , SpaCy, Nltk, AllenNLP, Linux , Slurm , AWS, GCP, Docker, Git , React, Redux, Webpack, Flask, Visual Studio Code , Vim , Blender, Photoshop, and Figma.
Other Interests	Analysis, algebra, topology, incentive theory, economics, cognitive science, neuroscience, videography, scuba diving, rock climbing, and fitness.