Shibani Santurkar

32 Vassar Street, 32G-818
Cambridge, MA 02139
✉ *shibani@mit.edu*
🖱 *people.csail.mit.edu/shibani/*

Education

2015–present **Ph.D. in Computer Science,** *Massachusetts Institute of Technology*
Research Advisors: Aleksander Mądry & Nir Shavit

2015–2017 **SM in Computer Science,** *Massachusetts Institute of Technology*
Advisor: Nir Shavit
Thesis: “Towards Generative Compression”

2010–2015 **Dual Degree (B.Tech and M.Tech),** *Indian Institute of Technology Bombay*
Electrical Engineering (Major) and Computer Science (Minor)

Research Interests

The focus of my research is on building a machine learning (ML) toolkit that allows for the reliable, robust, and auditable deployment of models in the real world. Specifically, my work revolves around:

- **Understanding current deep learning practices:** how various design choices (e.g., architectural components, datasets, and loss functions) impact model behavior in practice.
- **Studying generalization beyond training conditions:** characterizing and alleviating failures of models due to natural and adversarial shifts in the data distribution during deployment.
- **Building tools for fair and interpretable ML:** developing a fine-grained understanding of the features that models base predictions on, so as to identify model biases and possible ways to alleviate them.

Awards

2019-2021 **Google PhD Fellowship in Machine Learning**

2015 **Undergraduate Research Award**
Awarded for exceptional research at IIT Bombay.

2010-2014 **Dhirubhai Ambani Scholarship**
National higher education scholarship awarded by Dhirubhai Ambani foundation, India.

Selected Conference Publications

(* denotes equal contribution)

ICML 2020 **From ImageNet to Image Classification: Contextualizing Progress on Benchmarks**
D. Tsipras*, S. Santurkar*, L. Engstrom, A. Ilyas & A. Mądry

ICML 2020 **Identifying Statistical Bias in Dataset Replication**
L. Engstrom*, A. Ilyas*, S. Santurkar, D. Tsipras & A. Mądry

ICLR 2020 **Implementation Matters in Deep RL: A Case Study on PPO and TRPO**
L. Engstrom*, A. Ilyas*, S. Santurkar, D. Tsipras, F. Janoos, L. Rudolph & A. Mądry
Oral presentation

ICLR 2020 **A Closer Look at Deep Policy Gradients**
A. Ilyas*, L. Engstrom*, S. Santurkar, D. Tsipras, F. Janoos, L. Rudolph & A. Mądry
Oral presentation

NeurIPS 2019 **Image Synthesis with a Single (Robust) Classifier**
S. Santurkar*, D. Tsipras*, B.Tran*, A. Ilyas*, L. Engstrom* & A. Mądry

NeurIPS 2019 **Adversarial Examples Are Not Bugs, They Are Features**
A. Ilyas*, S. Santurkar*, D. Tsipras*, L. Engstrom*, B.Tran & A. Mądry
Spotlight presentation

ICLR 2019 **Robustness May be at Odds with Accuracy**
D. Tsipras*, S. Santurkar*, L. Engstrom*, A. Turner & A. Mądry

NeurIPS 2018 **How Does Batch Normalization Help Optimization?**
S. Santurkar*, D. Tsipras*, A. Ilyas* & A. Mądry
Oral presentation

NeurIPS 2018 **Adversarially Robust Generalization Requires More Data**
L. Schmidt, S. Santurkar, D. Tsipras, K. Talwar & A. Mądry
Spotlight presentation

ICML 2018 **A Classification–Based Study of Covariate Shift in GAN Distributions**
S. Santurkar, L. Schmidt & A. Mądry

ICML 2017 **Deep Tensor Convolution on Multicores**
D. Budden, A. Matveev, S. Santurkar, S. R. Chaudhuri & N. Shavit

Preprints

(* denotes equal contribution)

2020 **BREEDS: Benchmarks for Subpopulation Shift**
S. Santurkar*, D. Tsipras* & A. Mądry
arxiv:1906.00945

2019 **Adversarial Robustness as a Prior for Learned Representations**
L. Engstrom*, A. Ilyas*, S. Santurkar*, D. Tsipras*, B.Tran* & A. Mądry
arxiv:1906.00945

Work experience

6/2018–8/2018 **Google Inc.** *Intern*
Mentor: Ilya Mironov
Designed a general-purpose, configuration-free approach for differentially private data synthesis.

5/2017–8/2017 **Vicarious** *Intern*
Mentor: Huayan Wang
Create deep learning-based model for single image-based pose-prediction for robotic grasp planning.

Professional Service

2020 **"Towards Trustworthy ML" ICLR Workshop** *Co-organizer.*

2018-2021 **NeurIPS, ICML, ICLR** *Reviewer.*

2018 **The Quest Symposium on Robust, Interpretable AI (MIT)** *Co-organizer.*

2017-2020 **MIT UROP Mentor** *Supervised MIT undergraduate students on research projects.*

Teaching

Fall 2017 **6.336 Introduction to Numerical Simulation (MIT)** *Teaching Assistant*

Spring 2015 **EE739 Processor Design (IIT Bombay)** *Teaching Assistant*

Fall 2014 **EE301 Electromagnetic Waves (IIT Bombay)** *Teaching Assistant*