Shreya Shankar

San Francisco, CA
✉ *shreyashankar@berkeley.edu*
⌂ *www.shreya-shankar.com*
🐦 *sh_reya*
ⓖ *shreyashankar*
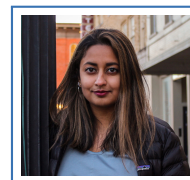
Education

Aug 2021–present
University of California, Berkeley, Berkeley, CA.
Ph.D. in Electrical Engineering and Computer Sciences (Databases)
Advised by Aditya Parameswaran

Sep 2015–Dec 2020
Stanford University, Stanford, CA.
M.S. in Computer Science (Artificial Intelligence), done part-time
B.S. in Computer Science (Systems)
Advised by Pat Hanrahan

Experience

Industry

May 2022–present
Research Engineer, *Meta*, Menlo Park, CA.
Researching and building an automatic data validation system for ML pipeline monitoring.

March 2021–Aug 2021
Entrepreneur in Residence, *Amplify Partners*, Menlo Park, CA.
Built open-source tools for machine learning software development (MLOps). Press release here.

June 2019–Jan 2021
Machine Learning Engineer, *Viaduct*, Palo Alto, CA.
Built systems and machine learning methods for large-scale time series data as the first ML engineer.
Worked with Airflow, Spark, SQL, Python, TensorFlow 2.0, XGBoost, Spark MLlib, and more.

Sep 2017–April 2019
Research Intern, *Google Brain*, Mountain View, CA.
Researched machine learning security and adversarial examples in collaboration with Stanford AI Lab.
Worked with TensorFlow 1.0, Python, and Borg.
Advised by Alex Kurakin and Ian Goodfellow.

June 2017–Sep 2017
Software Engineering Intern, *Facebook*, New York, NY.
Worked on Facebook's civic engagement team to connect users to their government representatives.
Worked with Hack (PHP), ReactJS, SQL, and Python.

Teaching

April 2020-June 2020
Teaching Assistant, *Stanford University*, Stanford, CA.
Served as a TA part-time for a remote version of CS110 (Principles of Computer Systems). Taught weekly sections and held weekly office hours via Zoom.

June 2018–Dec 2018
Head Teaching Assistant, *Stanford University*, Stanford, CA.
Served as head TA for CS106B (Programming Abstractions) and CS101 (Introduction to Computing Principles). Held weekly office hours. Helped write exams and homework grading criteria. Coordinated a staff of undergraduate section leaders.

Jan 2016–April 2018	**Undergraduate Section Leader**, *Stanford University*, Stanford, CA. Taught weekly sections for CS106A (Programming Methodologies) and CS106B (Programming Abstractions). Held weekly office hours. Graded assignments and exams.

Honors and Awards

- 2022 NDSEG Fellowship Recipient
- 2022 Hertz Foundation Fellowship Finalist
- 2022 P.D. Soros Fellowship Finalist
- 2021 UC Berkeley EECS Excellence Award Recipient
- 2020 Interact Fellowship Recipient
- 2015-2019 Rella Lou Danenberg Aldrich Scholarship Recipient
- 2016 Anita Borg Grace Hopper Scholarship Recipient
- 2016 Palantir Women in Technology Scholarship Recipient

Invited Talks

July 2022	**AI Explained Podcast**, *Fiddler.ai*. Discussed research challenges in ML monitoring.
July 2022	**O'Reilly Practical Tutorial**, *O'Reilly*. Gave a tutorial on monitoring ML systems.
June 2022	**AI + ML Meetup**, *Scale AI*, San Francisco, CA. Gave a talk on monitoring ML systems with feedback delays.
May 2022	**LinkedIn Live Podcast**, *FICO*. Discussed real-time monitoring of ML systems.
March 2022	**Observe Summit**, *Arize AI*. Gave a talk on monitoring ML without access to real-time labels.
March 2022	**Medical AI Lab Seminar**, *Harvard University*. Gave a talk on research challenges in ML monitoring.
February 2022	**MLCON 2.0**, *cnvrg.io*. Gave a talk on monitoring ML without access to real-time labels.
February 2022	**CS329S**, Stanford University, Stanford, CA. Gave a guest lecture on detecting distribution shift in data streams.
January 2022	**ML Reading Group**, *Twitter*, San Francisco, CA. Discussed observability and monitoring for ML systems.
January 2022	**MLOps Community Podcast**, *mlops.community*. Discussed CI/CD in ML pipelines.
January 2022	**CIDR 2022**, Chaminade, CA. Presenting Towards Observability for Machine Learning Pipelines.
December 2021	**AICamp**, *Google DevFest 2021*. Gave a talk on observability for ML systems and research directions in the field.
November 2021	**Toronto Machine Learning Virtual Summit**, *Toronto ML Society*, Toronto, Canada. Gave a talk on observability for ML systems and tutorial on building a ML pipeline with testing and monitoring.

November 2021	**RISECamp**, *UC Berkeley*, Berkeley, CA.
	Gave a talk on observability for ML systems and tutorial on building a ML pipeline with testing and monitoring.
October 2021	**Data Observability Summit**, *Facebook*, Menlo Park, CA.
	Gave a talk on observability for ML systems.
June 2021	**D&I Round Table**, *ACM SIGMOD/PODS Conference*.
	Participated in a panel on imposter syndrome.
June 2021	**MLOps World Conference**, *MLOps World*, Toronto, Canada.
	Gave a talk on debugging ML in production and demo-ed my open-source tracing tool.
May 2021	**Data + AI Summit**, *Databricks*.
	Gave a talk on debugging ML in production and demo-ed my open-source tracing tool.
March 2021	**MLOps Salon**, *Verta.AI*.
	Gave a talk on debugging ML in production and participated in a follow-up panel.
March 2021	**Practical AI Show**, *Clubhouse App*.
	Featured as a guest to discuss my recent retrospective on predictive modeling.
February 2021	**MLSys Seminar**, *Stanford University*, Stanford, CA.
	Gave a talk on debugging ML in production. Code and slides on my Github.
February 2021	**DSC102**, *University of California, San Diego*, San Diego, CA.
	Gave a talk on debugging ML in production. Code and slides on my Github.
February 2021	**Time Horizons Podcast**.
	Machine learning in industry.
February 2021	**NLP Zurich Meetup**, Zurich, Switzerland.
	Gave a talk on debugging ML in production. Code and slides on my Github.
January 2021	**OSCON**, O'Reilly.
	Participated as a panelist to discuss open source and machine learning.
January 2021	**CS329S**, Stanford University, Stanford, CA.
	Gave a tutorial on PyTorch and distributed training.
October 2020	**Machine Learning Podcast**.
	A day in the life on an Applied ML Researcher.
October 2020	**Data Engineered Podcast**.
	Lessons learned after a year of putting ML into production.
October 2020	**Datacast Podcast**.
	Computer Systems, Machine Learning Security Research, and Women in Tech.

Software

mltrace	This project enables coarse-grained lineage and tracing in complex data pipelines. *400+ stars.*
Toy ML Pipeline	This is a toy example of a standalone ML pipeline written entirely in Python. No external tools are incorporated into the master branch. I built it mainly to experiment with my ideas for ML tooling. *150+ stars.*
Create ML App	This project makes it easier to spin up a machine learning project locally in Python and handle various package dependencies using a Makefile. It abstracts away pip installs and virtual environment commands from the user. *500+ stars.*
GPT3 Sandbox	This project enables users to create cool web demos using OpenAI's GPT-3 API with just a few lines of Python. Co-authored with Bora Uyumazturk. *2.5k+ stars.*

Service

- Board member of SHE++, a 501(c)(3) nonprofit that improves diversity in tech.
- Former co-director of SHE++, a 501(c)(3) nonprofit that improves diversity in tech.
- Former financial officer of Stanford WiCS (Women in Computer Science).

Advising

Current

- Boyuan Deng (Undergrad, UC Berkeley)
- Yujie Wang (Undergrad, UC Berkeley)

Past

- Aditi Mahajan (Undergrad, UC Berkeley)
- Peter Maldonado (Undergrad, Stanford)

Reviewing

- ICLR 2022
- NeurIPS 2021
- ICML 2019 Workshop in Adversarial Machine Learning in Real-World Computer Vision Systems
- ICML 2019 Workshop in Security and Privacy of Machine Learning
- NeurIPS 2018 Workshop on Security in Machine Learning

Preprints and Publications

[1] S. Shankar and A. G. Parameswaran. Towards observability for machine learning pipelines. In *CIDR*, 2022.

[2] S. Shankar, B. Herman, and A. G. P. Parameswaran. Rethinking streaming machine learning evaluation. In *ICLR 2022 workshop: ML Evaluation Standards*, 2022.

[3] S. Dathathri, K. Dvijotham, A. Kurakin, A. Raghunathan, J. Uesato, R. R. Bunel, S. Shankar, J. Steinhardt, I. Goodfellow, P. S. Liang, and P. Kohli. Enabling certification of verification-agnostic networks via memory-efficient semidefinite programming. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 5318–5331. Curran Associates, Inc., 2020.

[4] G. F. Elsayed, S. Shankar, B. Cheung, N. Papernot, A. Kurakin, I. Goodfellow, and J. Sohl-Dickstein. Adversarial examples influence human visual perception. *Journal of Vision*, 19(10):190c–190c, Sep 2019.

[5] G. F. Elsayed, S. Shankar, B. Cheung, N. Papernot, A. Kurakin, I. Goodfellow, and J. Sohl-Dickstein. Adversarial examples that fool both computer vision and time-limited humans. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NeurIPS'18, page 3914–3924. Curran Associates, Inc., 2018.

[6] S. Shankar, Y. Halpern, E. Breck, J. Atwood, J. Wilson, and D. Sculley. No classification without representation: Assessing geodiversity issues in open data sets for the developing world. In *NIPS 2017 workshop: Machine Learning for the Developing World*, 2017.

Interests

Triathlons	Competed for Stanford's Triathlon team. Completed 2021 Ironman 70.3 Santa Cruz.
Hobbyist musician	Took classical piano and violin lessons from 2003-2015. Gave a senior recital in 2015. Now mainly playing pop songs and random Chopin works.
Writing	Member of a weekly writer's group in San Francisco. Technical writing available at personal website.
Intentional communities	Member of Phoenix House and Haight Street Commons, a network of co-ops in the Bay Area.