

Shreya Shankar

✉ shreyashankar@berkeley.edu

🌐 www.sh-reya.com

🐦 [sh_reya](https://twitter.com/sh_reya)

🌐 [shreyashankar](https://github.com/shreyashankar)



Education

- Aug **University of California, Berkeley**, Berkeley, CA.
2021–present Ph.D. in Electrical Engineering and Computer Sciences (Databases)
Advised by Aditya G. Parameswaran
- Sep **Stanford University**, Stanford, CA.
2015–Dec M.S. in Computer Science (Artificial Intelligence), done part-time
2020 B.S. in Computer Science (Systems)
Advised by Pat Hanrahan

Experience

Industry

- May **Research Engineer**, *Meta*, Menlo Park, CA.
2022–Nov Researched and engineered an automatic data validation system for ML pipeline monitoring.
2022 Published a paper at CIKM 2023.
- March **Entrepreneur in Residence**, *Amplify Partners*, Menlo Park, CA.
2021–Aug Built open-source tools for machine learning software development (MLOps). Press release here.
2021
- June **Machine Learning Engineer**, *Viaduct*, Palo Alto, CA.
2019–Jan Built systems and machine learning methods for large-scale time series data as the first ML
2021 engineer.
Worked with Airflow, Spark, SQL, Python, TensorFlow 2.0, XGBoost, Spark MLlib, and more.
- Sep **Research Intern**, *Google Brain*, Mountain View, CA.
2017–April Researched machine learning security and adversarial examples in collaboration with Stanford AI
2019 Lab.
Worked with TensorFlow 1.0, Python, and Borg.
Advised by Alex Kurakin and Ian Goodfellow.
- June **Software Engineering Intern**, *Facebook*, New York, NY.
2017–Sep Worked on Facebook's civic engagement team to connect users to their government representatives.
2017 Worked with Hack (PHP), ReactJS, SQL, and Python.

Teaching

- August **Graduate Student Instructor**, *UC Berkeley*, Berkeley, CA.
2022–Dec Served as a TA for DATA 101, a course on data engineering fundamentals. Held weekly office
2022 hours, taught weekly sections, and wrote some homeworks.
- April **Teaching Assistant**, *Stanford University*, Stanford, CA.
2020–June Served as a TA part-time for a remote version of CS110 (Principles of Computer Systems). Taught
2020 weekly sections and held weekly office hours via Zoom.

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

Preprints and Publications

- [1] S. Shankar, H. Li, P. Asawa, M. Hulsebos, Y. Lin, J. D. Zamfirescu-Pereira, H. Chase, W. Fu-Hinthorn, A. G. Parameswaran, and E. Wu. Spade: Synthesizing assertions for large language model pipelines, 2024, 2401.03038.
- [2] S. Shankar, L. Fawaz, K. Gyllstrom, and A. Parameswaran. Automatic and precise data validation for machine learning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM '23*, page 2198–2207, New York, NY, USA, 2023. Association for Computing Machinery.
- [3] A. G. Parameswaran, S. Shankar, P. Asawa, N. Jain, and Y. Wang. Revisiting prompt engineering via declarative crowdsourcing. *arXiv preprint arXiv:2308.03854*, 2023.
- [4] S. Shankar and A. Parameswaran. Towards observability for production machine learning pipelines. *Proceedings of the VLDB Endowment*, 15(13):4015–4022, 2022.
- [5] S. Shankar, S. Macke, A. Chasins, A. Head, and A. Parameswaran. Bolt-on, compact, and rapid program slicing for notebooks. *Proceedings of the VLDB Endowment*, 15(13):4038–4047, 2022.
- [6] S. Shankar, B. Herman, and A. G. P. Parameswaran. Rethinking streaming machine learning evaluation. In *ICLR 2022 workshop: ML Evaluation Standards*, 2022.
- [7] S. Shankar, R. Garcia, J. M. Hellerstein, and A. G. Parameswaran. Operationalizing machine learning: An interview study. *arXiv preprint arXiv:2209.09125*, 2022.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.

Honors and Awards

- 2023 Heidelberg Laureate Forum Young Scholar
- 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
- 2016 Anita Borg Grace Hopper Scholarship Recipient
- 2016 Palantir Women in Technology Scholarship Recipient

Selected Invited Talks

- January 2024 **Amazon Research**, *Amazon*.
Giving a talk on the SPADE project (assertions for LLM pipelines).
- November 2023 **INDE Lab**, *University of Amsterdam*.
Discussed my research on tooling for ML engineers.
- March 2023 **Core Data Tech Talk**, *Google*.
Discussed my research on ML observability.
- December 2022 **Normconf**, *Normconf*.
Gave a normie talk on data management for ML.
- October 2022 **Gradient Dissent Podcast**, *Weights & Biases*.
Discussed our interview study of ML engineers.
- February 2022 **CS329S**, *Stanford University, Stanford, CA*.
Gave a guest lecture on detecting distribution shift in data streams.
- 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.
- October 2021 **Data Observability Summit**, *Facebook, Menlo Park, CA*.
Gave a talk on observability for ML systems.
- 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.

Software

- Motion This research project provides a framework for building continually-updating ML applications.
- GATE This research project enables automatic and precise data validation for machine learning.
- 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

- On the student admissions committee for the UC Berkeley EECS PhD Program.
- Co-organizer of DEEM, a workshop at SIGMOD on end-to-end data management for machine learning.
- 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).

Mentoring

Current

- Parth Asawa (2 semesters + summer, CRA Undergraduate Award Honorable Mention)
- Rachel Lin (1 semester)

Past

- Boyuan Deng (2 semesters)
- Aditi Mahajan (2 semesters)
- Peter Maldonado (Summer)
- Yujie Wang (3 semesters + summer)

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

Interests

- Triathlons Competed for Stanford's Triathlon team. Completed 2021 Ironman 70.3 Santa Cruz.
- Writing Member of a weekly writer's group in San Francisco. Technical writing available at personal website.
- Intentional communities Alum of Phoenix House and Haight Street Commons, a network of co-ops in the Bay Area.