Osbert Bastani

▬▬ Employment

2018- **Research Assistant Professor in Computer and Information Sciences**, *University of Pennsylvania*, Philadelphia, PA.

2017-2018 **Postdoctoral Fellow in CSAIL**, *Massachusetts Institute of Technology*, Cambridge, MA.

2015 **Research Intern**, *Microsoft Research*, Cambridge, UK.

2014 **Research Intern**, *Google Research*, Mountain View, CA.

2013 **Research Intern**, *Technicolor Research Labs*, Palo Alto, CA.

▬▬ Education

2012-2018 **Ph.D. in Computer Science**, *Stanford University*, Stanford, CA.

2008-2012 **A.B. in Mathematics**, *Harvard University*, Cambridge, MA.

▬▬ Publications

Sangdon Park, Edgar Dobriban, Insup Lee, and Osbert Bastani. PAC prediction sets under covariate shift. *ICLR*, 2022.

Jason Y. Ma, Andrew Shen, Osbert Bastani, and Dinesh Jayaraman. Conservative and adaptive penalty for model-based safe reinforcement learning. *AAAI*, 2022.

Jason Y. Ma, Dinesh Jayaraman, and Osbert Bastani. Conservative offline distributional reinforcement learning. *NeurIPS*, 2021.

Yichen Yang, Jeevana P. Inala, Osbert Bastani, Yewen Pu, Armando Solar-Lezama, and Martin Rinard. Program synthesis guided reinforcement learning for partially observed environments. *NeurIPS (Spotlight)*, 2021.

Kishor Jothimurugan, Suguman Bansal, Osbert Bastani, and Rajeev Alur. Compositional reinforcement learning from logical specifications. *NeurIPS*, 2021.

Alexis Ross, Himabindu Lakkaraju, and Osbert Bastani. Learning models for actionable recourse. *NeurIPS*, 2021.

Soham Dan, Osbert Bastani, and Dan Roth. Few-shot novel concept learning for semantic parsing. *Findings of EMNLP*, 2021.

Jason Y. Ma, Jeevana I. Priya, Dinesh Jayaraman, and Osbert Bastani. Likelihood-based diverse sampling for trajectory forecasting. *ICCV*, 2021.

Favyen Bastani, Songtao He, Ziwen Jiang, Osbert Bastani, and Sam Madden. Skyquery: An aerial drone video sensing platform. *Onward*, 2021.

Radoslav Ivanov, Kishor Jothimurugan, Steve Hsu, Vaidya Shaan, Rajeev Alur, and Osbert Bastani. Compositional learning and verification of neural network controllers. *EMSOFT*, 2021.

Osbert Bastani, Shuo Li, and Anton Xue. Safe reinforcement learning via statistical model predictive shielding. *RSS*, 2021.

Kan Xu, Xuanyi Zhao, Hamsa Bastani, and Osbert Bastani. Group-sparse matrix factorization for transfer learning of word embeddings. *ICML*, 2021.

Jocelyn Chen, Aaron Lamoreaux, Xinyu Wang, Greg Durrett, Osbert Bastani, and Isil Dillig. Web question answering with neurosymbolic program synthesis. *PLDI*, 2021.

Osbert Bastani. Safe reinforcement learning with nonlinear dynamics via model predictive shielding. *ACC*, 2021.

Kishor Jothimurugan, Osbert Bastani, and Rajeev Alur. Abstract value iteration for hierarchical deep reinforcement learning. *AISTATS*, 2021.

Min Wen, Osbert Bastani, and Ufuk Topcu. Algorithms for fairness in sequential decision making. *AISTATS*, 2021.

Sangdon Park, Shuo Li, Insup Lee, and Osbert Bastani. PAC confidence predictions for deep neural network classifiers. *ICLR*, 2021.

Jeevana P. Inala, Yichen Yang, James Paulos, Yewen Pu, Osbert Bastani, Vijay Kumar, Martin Rinard, and Armando Solar-Lezama. Neurosymbolic transformers for multi-agent communication. *NeurIPS*, 2020.

Jiani Huang, Calvin Smith, Osbert Bastani, Rishabh Singh, Aws Albarghouthi, and Mayur Naik. Generating programmatic referring expressions via program synthesis. *ICML*, 2020.

Himabindu Lakkaraju, Nino Arsov, and Osbert Bastani. Robust and stable black box explanations. *ICML*, 2020.

Yanju Chen, Chenglong Wang, Osbert Bastani, Isil Dillig, and Yu Feng. Program synthesis using deduction-guided reinforcement learning. *CAV*, 2020.

Shuo Li and Osbert Bastani. Robust model predictive shielding for safe reinforcement learning with stochastic dynamics. *ICRA*, 2020.

Osbert Bastani. Sample complexity of estimating the policy gradient for nearly deterministic dynamical systems. *AISTATS*, 2020.

Sangdon Park, Osbert Bastani, Jim Weimer, and Insup Lee. Calibrated prediction with covariate shift via unsupervised domain adaptation. *AISTATS*, 2020.

Sangdon Park, Osbert Bastani, Nikolai Matni, and Insup Lee. PAC confidence sets for deep neural networks via calibrated prediction. *ICLR*, 2020.

Jeevana P. Inala, Osbert Bastani, Zenna Tavares, and Armando Solar-Lezama. Synthesizing programmatic policies that inductively generalize. *ICLR*, 2020.

Himabindu Lakkaraju and Osbert Bastani. "How do I fool you?": Manipulating user trust via misleading black box explanations. *AIES*, 2020.

Kishor Jothimurugan, Rajeev Alur, and Osbert Bastani. Composable specifications for reinforcement learning. *NeurIPS*, 2019.

Osbert Bastani, Xin Zhang, and Armando Solar-Lezama. Verifying fairness properties via concentration. *OOPSLA*, 2019.

Jai Chen, Jiayi Wei, Yu Feng, Osbert Bastani, and Isil Dillig. Relational verification using reinforcement learning. *OOPSLA*, 2019.

Zhengkai We, Evan Johnson, Wei Yang, Osbert Bastani, Dawn Song, Jian Peng, and Tao Xie. Reinam: Reinforcement learning for input-grammar inference. *FSE*, 2019.

Arbaaz Khan, Chi Zhang, Shuo Li, Jiayue Wu, Brent Schlotfeldt, Sarah Tang, Alejandro Ribeiro, Osbert Bastani, and Vijay Kumar. Learning safe unlabeled multi-robot planning with motion constraints. *IROS*, 2019.

Halley Young, Osbert Bastani, and Mayur Naik. Learning neurosymbolic generative models via program synthesis. *ICML*, 2019.

Osbert Bastani, Yewen Pu, and Armando Solar-Lezama. Verifiable reinforcement learning via policy extraction. *NeurIPS*, 2018.

Osbert Bastani, Rahul Sharma, Alex Aiken, and Percy Liang. Active learning of points-to specifications. *PLDI*, 2018.

Yu Feng, Ruben Martins, Osbert Bastani, and Isil Dillig. Program synthesis using conflict-driven learning. *PLDI*, 2018.

Osbert Bastani, Carolyn Kim, and Hamsa Bastani. Interpretability via model extraction. *FAT/ML*, 2017.

Osbert Bastani, Rahul Sharma, Alex Aiken, and Percy Liang. Synthesizing program input grammars. *PLDI*, 2017.

Yu Feng, Osbert Bastani, Ruben Martins, Isil Dillig, and Saswat Anand. Automated synthesis of semantic malware signatures using maximum satisfiability. *NDSS*, 2017.

Osbert Bastani, Yani Ioannou, Lenonidas Lampropoulos, Dimitrios Vytiniotis, Aditya Nori, and Antonio Criminisi. Measuring neural net robustness with constraints. *NIPS*, 2016.

Lazaro Clapp, Osbert Bastani, Saswat Anand, and Alex Aiken. Minimizing GUI event traces. *FSE*, 2016.

Osbert Bastani, Saswat Anand, and Alex Aiken. An interactive approach to mobile app verification. *MobileDeLi*, 2015.

Osbert Bastani, Saswat Anand, and Alex Aiken. Interactively verifying absence of explicit information flows in android apps. *OOPSLA*, 2015.

Osbert Bastani, Saswat Anand, and Alex Aiken. Specification inference using context-free language reachability. *POPL*, 2015.

Osbert Bastani, Christopher Hillar, Dimitar Popov, and Maurice Rojas. Randomization, sums of squares, near-circuits, and faster real root counting. *Contemporary Mathematics*, 2011.

Awards & Honors

2018	**PLDI Distinguished Paper Award**.
2015-2017	**Google Ph.D. Fellowship**.
2012-2013	**Stanford School of Engineering Fellowship**.

Invited Talks

2021	**IFDS Summer Workshop on Statistical Approaches to Understanding Modern Machine Learning Methods**.
2021	**Simons Institute Workshop on Games and Equilibria**.
2020	**ICML Workshop on Explainable AI: Beyond Deep Models and Classifiers**.
2020	**UCSD AI Seminar**.
2020	**UCSB Programming Languages Seminar**.

2019	**UT Austin Programming Languages Seminar**.
2019	**ICLR Workshop on Debugging Machine Learning Models**.
2018	**CPS Week Workshop on Design and Analysis of Robust Systems**.
2017	**University of Pennsylvania Computer Science Seminar**.
2017	**Cornell University Computer Science Colloquium**.
2017	**Northwestern University Computer Science Seminar**.
2017	**Penn State University Computer Science Seminar**.

Funded Grants

2020-2025	**Co-PI**, *NSF Expeditions: Understanding the World Through Code*.
2020-2023	**Co-PI**, *ARO MURI: Robust Concept Learning and Lifelong Adaptation Against Adversarial Attacks*.
2019-2022	**PI**, *NSF SHF: Small: Inferring Specifications for Blackbox Code*.
2019-2022	**Co-PI**, *DARPA Learning with Less Labels*.

Ph.D. Student Advising

2021-	**Joey Velez-Ginorio (co-advised with Steve Zdancewic)**, *Current student*.
2020-	**Jason Ma (co-advised with Dinesh Jayaraman)**, *Current student*.
2020-	**Shuo Li (co-advised with Insup Lee)**, *Current student*.
2019-	**Halley Young**, *Current student*.
2019-	**Stephen Mell (co-advised with Steve Zdancewic)**, *Current student*.
2019-2021	**Sangdon Park (co-advised with Insup Lee)**, *Thesis: Uncertainty Estimation Toward Safe AI*.

Undergraduate/Masters Student Advising

2021-	**Aishwarya Wesanekar (masters)**.
2021-	**Utkarsh Kashyap (masters)**.
2020-	**Ryan Gannon (undergraduate)**.
2020-	**Aryan Singh (undergraduate)**.
2020-	**Jian Zhang (undergraduate)**.
2020-	**Sarah Luthra (undergraduate)**.
2020-2021	**George Tolkachev (masters)**.
2020-2021	**Wanqiao Xu (undergraduate)**.
2020	**Maxwell Du (undergraduate)**.
2019	**Brian Heath (masters)**.

Service

2018-2021	**ICML, NeurIPS, ICLR, AISTATS, AAAI, CVPR, IJRR, ICRA, R-AL, TAC, ACC**, Reviewer.
2022	**OOPSLA 2022**, Program Committee Member.
2021	**PLDI 2022**, Program Committee Member.
2020	**PLDI 2021**, Program Committee Member.
2019	**POPL 2020**, Program Committee Member.
2019	**CAV 2019**, Program Committee Member.
2018	**PLDI 2019**, External Program Committee Member.
2017	**PLDI 2018**, External Program Committee Member.

Teaching

2016	**Teaching Assistant**, *Stanford University*, CS 265: Randomized Algorithms and Probabilistic Analysis.
2016	**Teaching Assistant**, *Stanford University*, CS 229T: Statistical Learning Theory.
2011	**Teaching Assistant**, *Harvard University*, Math 124: Number Theory.