

Dorsa Sadigh

Contact	142 Gates Computer Science Building 353 Jane Stanford Way, Stanford, CA 94305 +1 (949) 929-3559 dorsa@cs.stanford.edu	https://dorsa.fyi https://github.com/Stanford-ILIAD
Research Interests	Robotics, Human-Robot Interaction, Control Theory, Formal Methods	
Current Position	Stanford University Assistant Professor Department of Computer Science and Department of Electrical Engineering	September 2017 - present
Education	University of California, Berkeley Ph.D. in Electrical Engineering and Computer Sciences Advisors: Sanjit Seshia and Shankar Sastry Thesis: <i>Safe and Interactive Autonomy: Control, Learning, and Verification</i>	2017
	University of California, Berkeley B.S. in Electrical Engineering and Computer Sciences	2012
Awards	<ul style="list-style-type: none">- National Science Foundation CAREER Award- Selected to give Gilbreth Lecture at National Academy of Engineering- Google Faculty Research Award- Amazon Research Award- Best Cognitive Robotics Paper Finalist, IROS.- "Active Information Gathering over Human Internal State"- Leon O. Chua Award for excellence in non-linear science, EECS Department, UC Berkeley- Google Anita Borg Scholarship- National Defense Science and Engineering Graduate Fellowship- National Science Foundation Graduate Research Fellowship- CRA Outstanding Undergraduate Researcher Award- Semiconductor Research Corporation (SRC-URO) Scholar- Arthur M. Hopkin Award for seriousness of purpose & high academic achievement, EECS Department, UC Berkeley- Eta Kappa Nu (EECS Honors Society)	<ul style="list-style-type: none">2020202020202019201620162016201320132012201120112010
Teaching	CS 237B: Robot Autonomy II Instructor, Stanford University.	Winter 2020
	CS 221: Artificial Intelligence Instructor, Stanford University.	Spring 2018 - 2019, Fall 2019

CS 521: Seminar on AI Safety **Spring 2018, 2020**
Instructor, Stanford University.

CS 333: Safe and Interactive Robotics **Fall 2017 - 2018**
Instructor, Stanford University.

EE120: Signals and Systems **Spring 2015**
TA for Murat Arcaç. UC Berkeley.

EECS20N: Structure & Interpretation of Systems & Signals **Spring 2014**
TA for David Tse and Thomas Courtade. UC Berkeley.

EECS20N: Structure & Interpretation of Systems & Signals **Spring 2011**
TA for Edward Lee and Pieter Abbeel. UC Berkeley.

Advising & Mentoring

Current Graduate Students

Erdem Bıyık, Minae Kwon, Malayandi Palaniappan (co-advised with Stephen Boyd), Mengxi Li (co-advised with Jeannette Bohg), Zhangjie Cao, Siddhartha Karamcheti (co-advised with Percy Liang)

Current Undergraduate Students

Gleb Shevchuck, Woody Wang, Kenneth Wang

Past Undergraduate Students

Nick Landolfi (Ph.D. student at Stanford), Zhiyang He (Ph.D. student at UC Berkeley)

Outreach

Stanford AI Mentorship Program

2018 - present

I have organized the Stanford AI mentorship program, where we connect underrepresented minorities and female undergraduate students interested in AI with Ph.D. students at Stanford to meet monthly and discuss research and career choices.

Faculty Mentor for Stanford Robotics Club

2017 - present

I mentor the Stanford undergraduate Robotics Club. Every year they work towards participating in a robotics competition. They have won the third place in the University Rover Challenge in 2019.

Faculty Mentor for Inclusion in AI

2018 - present

I mentor the Stanford AI Lab graduate group “Inclusion in AI”. The group holds regular social and networking events for Stanford AI Lab graduate students.

Talks at Women and Inclusion in STEM events and panels

AI4ALL summer program, Girls Who Code summer program, Gender in Robotics Workshop at Stanford, Berkeley-Stanford Meetup, Rising Stars (EECS) of 2018, Rising Stars (Mechanical Engineering) of 2019, Inclusion in AI.

Talks at Graduate and Undergraduate Student Groups

Undergrad CS Women (WiCS), Grad Engineering Women (SWE), SAIL (Stanford AI Lab women), Women in Electrical Engineering, Women in Aero/Astro, Fire-Side chat with Stanford Undergrads.

	EEGSA Outreach Member	2012 - 2017
	Visiting local K-12 schools and presenting engineering projects and demonstrations.	
	WICSE Outreach Coordinator	2014 - 2015
	Organizing events and outreach activities aiming young girls involvements in STEM.	
Work Experience	Microsoft Research, Redmond	June - August 2015
	Internship at the Adaptive Systems and Interaction group with Ashish Kapoor and Eric Horvitz.	
	Stanford Research Institute, International	June - August 2013
	Internship at the Computer Science Laboratory in the formal methods group with Ashish Tiwari.	
Professional Activities	Center for AI Safety at Stanford	2018 - present
	Founding member of the Center for AI Safety at Stanford along with Mykel Kochenderfer, Clark Barrett, and David Dill. The center is focused on safety and verification issues for AI and machine learning systems.	
	Human-Centered AI Institute	2018 - present
	Member of the design committee of Human-Centered AI Institute at Stanford.	
	Program Co-Chair	2018 - 2019
	Bay Area Robotics Symposium	
	Program Committee (Associate Editor)	
RSS 2020, HRI 2020, L4DC 2020, CAV 2019, HSCC 2019, CoRL 2018, ICRA 2018, HSCC Repeatability Eval 2016.		
	Workshop Organizer	
	Reliable Autonomy for Human-Cyber-Physical Systems at NSF PI meeting , Virginia, 2018.	
	Workshop Organizer	
	IROS 2016: Perspectives on Analysis and Design of Human-Centered Robotics .	
	IROS 2019: Learning Representations for Planning and Control.	
	DREAM Seminar Organizer	2015 - 2017
	https://embedded.eecs.berkeley.edu/seminar/	
	External Reviewer for Conferences, Journals, and Grant Panels	
	- <i>Robotics</i> : RSS 2017-19, CoRL 2018-19, WAFR 2016, ICRA 2017-18, HRI 2018, TASE 2015, ACM TECS 2015	
	- <i>Control Theory</i> : HSCC 2013-2016, CDC 2015-2016, ACC 2016-2017, TCST 2017	
	- <i>Formal Methods</i> : CAV 2016, FM 2015, HVC 2016, VMCAI 2015	
	- <i>NSF Proposal Panels</i>	
Invited Talks	Keynote at HSCC.	
	Beyond Theory of Mind: Learning & Influencing Conventions	2020

IPAM workshop on Intersections between Control, Learning, and Optimization. – “–.	2020
University of Washington, Robotics Colloquium. – “–.	2020
Gilbreth Lecture, National Academy of Engineering Influencing Interactions in Autonomous Driving	2020
Keynote at FMCAD. A journey about Safety of Autonomous Systems	2019
Frontiers of Engineering, National Academy of Engineering. Influencing Interactions in Autonomous Driving	2019
RSS 2019 Workshop on Safe Autonomy. –“–.	2019
Learning for Dynamics and Control Workshop. Influencing Interactive Mixed-Autonomy Systems.	2019
ICML Workshop on AI for Autonomous Driving. –“–.	2019
MIT, Department Seminar. Interactive Autonomy: Learning and Control for Human-Robot Systems.	2019
University of Washington, Department Seminar. –“–.	2019
Cornell, Department Seminar. –“–.	2019
CalTech, IST Seminar. –“–.	2019
USC, CPS Seminar. –“–.	2019
University of Maryland, Robotics Seminar. –“–.	2019
Theoretical Machine Learning Simons Foundation Workshop. –“–.	2019
Schloss Dagstuhl on Verification and Synthesis for Human-Robot Interaction. Reward Functions and Specifications	2019
NeurIPS workshop on Imitation Learning and its Challenges in Robotics. Active Learning of Humans’ Preferences.	2018
UAI workshop on Safety, Risk and Uncertainty in RL. –“–.	2018
UC Berkeley, Center for Human Compatible AI. –“–.	2018
NeurIPS workshop on Machine Learning for Intelligent Transportation Systems. Beating Congestion using Autonomous Cars.	2018
Halmstad University. Reactive Synthesis and Human Modeling for Human-Robot Systems.	2018

University of Washington, Robotics Seminar. Safe and Interactive Robotics. 2018	
UC Santa Barbara, Robotics Seminar. –“–.	2018
UC Santa Cruz, Robotics Seminar. –“–.	2018
Chinese University of Hong Kong in Shenzhen. –“–.	2018
Stanford University, Department Seminar. Towards a Theory of Safe and Interactive Autonomy.	2017
MIT, Department Seminar. –“–.	2017
UC Berkeley, Department Seminar. –“–.	2017
CMU, Department Seminar. –“–.	2017
Princeton, Department Seminar. –“–.	2017
USC, Department Seminar. –“–.	2017
Cornell, Department Seminar. –“–.	2017
UC San Diego, Department Seminar. –“–.	2017
UC Los Angeles, Department Seminar. –“–.	2017
University of Michigan, Department Seminar. –“–.	2017
UT Austin, Department Seminar. –“–.	2017
Georgia Tech, Department Seminar. –“–.	2017
University of Pennsylvania, Department Seminar. –“–.	2017
Schloss Dagstuhl on Machine Learning and Formal Methods. Planning for Cars that Coordinate with People.	2017
Schloss Dagstuhl on Non-Zero-Sum-Games and Control. Correctness and Control for Human-Cyber-Physical Systems.	2015
Microsoft Research, Redmond. Controller Synthesis for Human-in-the-Loop Systems	2014

**Conference &
Journal
Publications**

- [42] Jonathan Mern, Dorsa Sadigh, Mykel Kochenderfer. Object Exchangability in Reinforcement Learning. *2020 American Control Conference (ACC), July 2020.*
- [41] Minae Kwon, Erdem Bıyık, Aditi Talati, Karan Bhasin, Dylan P. Losey, Dorsa Sadigh. When Humans Aren’t Optimal: Robots that Collaborate with Risk-Aware Humans. *ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2020.*
- [40] Yuhang Che, Allison M. Okamura, Dorsa Sadigh. Efficient and Trustworthy

Social Navigation Via Explicit and Implicit Robot-Human Communication. *IEEE Transactions on Robotics (TRO)*, 2019.

[39] Dylan P. Losey, Krishnan Srinivasan, Ajay Mandlekar, Animesh Garg, Dorsa Sadigh. Controlling Assistive Robots with Learned Latent Actions. *International Conference on Robotics and Automation (ICRA)*, May 2020.

[38] Dylan P. Losey, Mengxi Li, Jeannette Bohg, Dorsa Sadigh. Learning from My Partner’s Actions: Roles in Decentralized Robot Teams. *Conference on Robot Learning (CoRL)*, 2019.

[37] Erdem Bıyık, Malayandi Palan, Nicholas Landolfi, Dylan P. Losey, Dorsa Sadigh. Asking Easy Questions: A User-Friendly Approach to Active Reward Learning. *Conference on Robot Learning (CoRL)*, 2019.

[36] Dylan P. Losey, Dorsa Sadigh. Robots that Take Advantage of Human Trust. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, November 2019.

[35] Chandrayee Basu, Erdem Bıyık, Zhixun He, Mukesh Singhal, Dorsa Sadigh. Active Learning of Reward Dynamics from Hierarchical Queries. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, November 2019.

[34] Erdem Bıyık, Daniel A. Lazar, Dorsa Sadigh, Ramtin Pedarsani. The Green Choice: Learning and Influencing Human Decisions on Shared Roads. *Proceedings of the 58th IEEE Conference on Decision and Control (CDC)*, December 2019.

[33] Minae Kwon, Mengxi Li, Alexandre Bucquet, Dorsa Sadigh. Influencing Leading and Following in Human-Robot Teams. *Robotics: Science and Systems (RSS)*, June 2019.

[32] Malayandi Palan, Gleb Shevchuk, Nicholas C. Landolfi, Dorsa Sadigh. Learning Reward Functions by Integrating Human Demonstrations and Preferences. *Robotics: Science and Systems (RSS)*, June 2019.

[31] Tianhe Yu, Gleb Shevchuk, Dorsa Sadigh, Chelsea Finn. Unsupervised Visuomotor Control through Distributional Planning Networks. *Robotics: Science and Systems (RSS)*, June 2019.

[30] Erdem Bıyık, Jonathan Margoliash, Shahrouz Ryan Alimo, Dorsa Sadigh. Efficient and Safe Exploration in Deterministic Markov Decision Processes with Unknown Transition Models. *2019 American Control Conference (ACC)*, July 2019.

[29] Elis Steffansson, Jaime Fisac, Dorsa Sadigh, Shankar Sastry, Karl H. Johansson. Human-Robot Interaction for Truck Platooning Using Hierarchical Dynamic Games. *European Control Conference (ECC)*, June 2019.

[28] Ashwini Pople, Roberto Martin-Martin, Patrick Goebel, Vincent Chow, Hans M. Ewald, Junwei Yang, Zenkai Wang, Amir Sadeghian, Dorsa Sadigh, Silvio Savarese, Marynel Vazquez. Deep Local Trajectory Planning and Control for Robot Navigation. *International Conference on Robotics and Automation (ICRA)*, May 2019.

- [27] Jaime F. Fisac, Eli Bronstein, Elis Stefnasson, Dorsa Sadigh, S. Shankar Sastry, Anca D. Dragan. Hierarchical Game-Theoretic Planning for Autonomous Vehicles. *International Conference on Robotics and Automation (ICRA)*, May 2019.
- [26] Erdem Bıyık, Dorsa Sadigh. Batch Active Preference-Based Learning of Reward Functions. *Conference on Robot Learning (CoRL)*, 2018.
- [25] Erdem Bıyık, Daniel A. Lazar, Ramtin Pedarsani, Dorsa Sadigh. Altruistic Autonomy: Beating Congestion on Shared Roads . *International Workshop on Algorithmic Foundations of Robotics (WAFR)*, 2018.
- [24] Daniel Lazar, Kabir Chandrasekher, Ramtin Pedarsani, Dorsa Sadigh. Maximizing Road Capacity Using Cars that Influence People. *IEEE Conference on Decision and Control (CDC)*, 2018.
- [23] Jiaming Song, Hongyu Ren, Dorsa Sadigh, Stefano Ermon. Multi-Agent Generative Adversarial Imitation Learning. *Conference on Neural Information Processing Systems (NeurIPS)*, 2018.
- [22] Dorsa Sadigh, S. Shankar Sastry, Sanjit Seshia. Verifying Robustness of Human-Aware Autonomous Cars . *IFAC conference on Cyber-Physical and Human Systems (CPHS)*, 2018.
- [21] Dorsa Sadigh, Nick Landolfi, S. Shankar Sastry, Sanjit A. Seshia, Anca Dragan. Planning for Autonomous Cars that Leverages Effects on Human Actions. *Invited to Autonomous Robots (AURO)*, 2018.
- [20] Susmit Jha, Vasumathi Raman, Dorsa Sadigh, Sanjit A. Seshia. Safe Autonomy Under Perception Uncertainty Using Chance-Constrained Temporal Logic . *Journal of Automatic Reasoning (JAR)*, 2018.
- [19] Dorsa Sadigh. Safe and Interactive Autonomy: Control, Learning, and Verification. *Ph.D. Dissertation. EECS Department, University of California, Berkeley, August 2017.*
- [18] Dorsa Sadigh, S. Shankar Sastry, Sanjit Seshia, Anca Dragan. Active Preference-Based Learning of Reward Functions. *Robotics: Science and Systems Conference (RSS)*, July 2017.
- [17] Negar Mehr, Dorsa Sadigh, Roberto Horowitz, S. Shankar Sastry, Sanjit Seshia. Stochastic Predictive Freeway Ramp Metering from Signal Temporal Logic Specifications. *American Control Conference (ACC)*, May 2017.
- [16] Dorsa Sadigh, S. Shankar Sastry, Sanjit Seshia, Anca Dragan. Information Gathering Actions over Human Internal State. *International Conference on Intelligent Robots and Systems (ICRA)*, 2016. **(best paper in cognitive robotics award finalist).**
- [15] Tara Rezvani, Katherine Driggs-Campbell, Dorsa Sadigh, S. Shankar Sastry, Sanjit Seshia, Ruzena Bajcsy. Towards Trustworthy Automation: User Interfaces that Convey Internal and External Awareness. *IEEE Intelligent Transportation Systems Conference (ITSC)*, November 2016.

- [14] Dorsa Sadigh, S. Shankar Sastry, Sanjit Seshia, Anca Dragan Planning for Autonomous Cars that Leverages Effects on Human Actions . *Robotics: Science and Systems Conference (RSS)*, 2016.
- [13] Dorsa Sadigh, Ashish Kapoor. Safe Control under Uncertainty with Probabilistic Signal Temporal Logic. *Robotics: Science and Systems Conference (RSS)*, 2016.
- [12] Shromona Ghosh, Dorsa Sadigh, Pierluigi Nuzzo, Vasumathi Raman, Alexandre Donze, Alberto Sangiovanni-Vincentelli, S. Shankar Sastry, Sanjit Seshia. Diagnosis and Repair for Synthesis from Signal Temporal Logic Specifications. *Conference on Hybrid Systems: Computation and Control (HSCC)*, 2016.
- [11] Sanjit A. Seshia, Dorsa Sadigh, S. Shankar Sastry. Formal Methods for Semiautonomous Driving. *Design and Automation Conference (DAC)*, 2015.
- [10] Vasumathi Raman, Alexandre Donze, Dorsa Sadigh, Richard M. Murray, Sanjit Seshia. Reactive Synthesis from Signal Temporal Logic Specifications. *Conference on Hybrid Systems: Computation and Control (HSCC)*, 2015.
- [9] Dorsa Sadigh, Eric S. Kim, Samuel Coogan, S. Shankar Sastry, Sanjit Seshia. A Learning Based Approach to Control Synthesis of Markov Decision Processes for Linear Temporal Logic Specifications. *IEEE Conference on Decision and Control (CDC)*, 2014.
- [8] Dorsa Sadigh, Henrik Ohlsson, S. Shankar Sastry, Sanjit Seshia. Robust Subspace System Identification via Weighted Nuclear Norm Optimization. *International Federation of Automatic Control (IFAC)*, 2014.
- [7] Dorsa Sadigh, Katherine Driggs-Campbell, Alberto Puggelli, Wenchao Li, Victor Shia, Ruzena Bajcsy, Alberto Sangiovanni-Vincentelli, Shankar Sastry, and Sanjit Seshia. Data-driven probabilistic modeling and verification of human driver behavior. *Formal Verification and Modeling in Human-Machine Systems (AAAI Spring Symposium)*, 2014.
- [6] Dorsa Sadigh, Katherine Driggs Campbell, Ruzena Bajcsy, S. Shankar Sastry, Sanjit Seshia. User Interface Design and Verification for Semi-autonomous Driving. *Conference on High Confidence Networked Systems*, 2014.
- [5] Ashish Tiwari, Bruno Dutertre, Dejan Jovanovic, Thomas de Candia, Dorsa Sadigh, Sanjit Seshia. Safety Envelop in Security. *Conference on High Confidence Networked Systems (HiCoNS)*, 2014.
- [4] Wenchao Li, Dorsa Sadigh, S. Shankar Sastry, Sanjit Seshia. Synthesis for Human-in-the-Loop Control Systems. *Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, 2014.
- [3] Dorsa Sadigh, Sanjit Seshia and Mona Gupta. Automating Exercise Generation: A Step towards Meeting the MOOC Challenge for Embedded Systems. *Workshop on Embedded Systems Education*, 2012.
- [2] Orna Kupferman, Dorsa Sadigh, and Sanjit A. Seshia. Synthesis with Clairvoyance. *Haifa Verification Conference (HVC)*, 2011.
- [1] Jonathan Kotker, Dorsa Sadigh, and Sanjit A. Seshia. Timing Analysis of Interrupt-

Driven Programs under Context Bounds. *Formal Methods in Computer Aided Design (FMCAD), 2011.*

Technical Reports & Workshop Papers

[4] Robert X. D. Hawkins, Minae Kwon, Dorsa Sadigh, Noah D. Goodman. Continual Adaptation for Efficient Machine Communication. *Proceedings of the ICML Workshop on Adaptive & Multitask Learning: Algorithms & Systems, June 2019. (best student paper).*

[3] Jiaming Song, Hongyu Ren, Dorsa Sadigh, Stefano Ermon. Multi-Agent Generative Adversarial Imitation Learning. *International Conference on Learning Representations (ICLR), Workshop Track, April 2018.*

[2] Sanjit Seshia, Dorsa Sadigh, S. Shankar Sastry. Towards Verified Artificial Intelligence. *Technical Report, July 2016.*

[1] Debadeepta Dey, Dorsa Sadigh, Ashish Kapoor. Fast Safe Mission Plans for Autonomous Vehicles. *Proceedings of Robotics: Science and Systems Workshop, June 2016.*

Dissertation

Dorsa Sadigh. Safe and Interactive Autonomy: Control, Learning, and Verification. *Ph.D. Dissertation; EECS Department, University of California, Berkeley, August 2017.*