

Andi Peng

MIT CSAIL (45.711)
andipeng@mit.edu
andipeng.com

Education

2023 -	Massachusetts Institute of Technology Ph.D., Electrical Engineering and Computer Science Advisors: Jacob Andreas and Julie Shah	Cambridge, MA
2023	Massachusetts Institute of Technology M.S., Electrical Engineering and Computer Science Advisors: Pulkit Agrawal and Julie Shah	Cambridge, MA
2018	Yale University , <i>cum laude, with distinction</i> B.S., Cognitive Science B.A., Global Affairs Awarded Douglas A. Beck Prize <i>for high academic achievement, leadership potential, personal integrity, and commitment to public service</i>	New Haven, CT

Research Positions

May 2024 -	Research Scientist, Anthropic	San Francisco, CA
Jan 2024 -	Special Government Employee (SGE), Defense Innovation Unit	Mountain View, CA
Fall 2023	Research Intern, Boston Dynamics AI Institute Host: Jessica Hodgins	Cambridge, MA
Summer 2023	Research Intern, MIT-IBM Watson AI Lab Host: Chuang Gan	Cambridge, MA
Summer-Fall 2021	Research Intern, Facebook AI Research (FAIR) Hosts: Aravind Rajeswaran and Vikash Kumar	Pittsburgh, PA
2018 - 2020	AI Resident, Microsoft Research Hosts: Ece Kamar, Besmira Nushi, Emre Kiciman, Kori Inkpen	Redmond, WA
2018	Policy Analyst, White House Office of Science and Technology Policy (OSTP) Research Associate, National Institute for Standards and Technology (NIST)	Washington, DC
Summer 2016	MARTI Researcher, NASA Glenn Research Center Hosts: Justin Gray and Jeffrey Chin	Cleveland, OH

Fellowships, Honors, and Awards

2024	Robotics: Science and Systems (RSS) Pioneers	
2023 - 2025	Open Philanthropy Research Grant	\$145,422
2020 - 2025	NSF Graduate Research Fellowship	\$138,000
2018	Fox Fellowship, University of Cambridge	\$30,000 (declined)
2017	Truman Scholarship	\$30,000
2017	Grand Strategy Research Grant, Yale University	\$4,000
2016	John D. Heinz Fellowship, Yale University	\$14,000
2015	Nathan Hale Scholarship, Yale University	\$55,000

	<i>A special distinction that reflects the university's esteem for past and future achievements</i>
2014	The President's Volunteer Service Award, Barack Obama's Council on Service and Civic Participation
2013	Appointment to the U.S. Military Academy at West Point (declined)
	<i>Nominated by Senator Sherrod Brown and Congressman Steve LaTourette</i>

Journal Publications

* equal contribution

Nature Machine Intelligence	The Future of Open Human Feedback Shachar Don-Yehiya, Ben Burtenshaw, ... Andi Peng , ..., Leshem Choshen. <i>Nature Machine Intelligence</i> , 2025
Preprint	Getting Aligned on Representational Alignment Ilia Sucholutsky*, Lukas Muttenthaler*, Adrian Weller, Andi Peng , ..., Thomas L. Griffiths. <i>Under review</i>
TMLR (Featured Certification)	Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback Stephen Casper*, Xander Davies*, ..., Andi Peng , ..., Dylan Hadfield-Menell. <i>Transactions on Machine Learning Research (TMLR)</i> , 2023
Nature	Make Greenhouse-Gas Accounting Reliable—Build Interoperable Systems Amy Luers, Leehi Yona, ..., Andi Peng , ..., Lucas Joppa. <i>Nature</i> , 2022

Conference Publications

ICLR 2025	Learning How Hard to Think: Input-Adaptive Allocation of LM Computation Mehul Damani, Idan Shenfeld, Andi Peng , Andreea Bobu, Jacob Andreas. <i>International Conference on Learning Representations (ICLR)</i> , 2025
CoRL 2024	Adaptive Language-Guided Abstraction from Contrastive Explanations Andi Peng , Belinda Z. Li, Ilia Sucholutsky, Nishanth Kumar, Julie A. Shah, Jacob Andreas, Andreea Bobu. <i>Conference on Robot Learning (CoRL)</i> , 2024
NeurIPS 2024	Constrained Human-AI: An Inclusive Embodied AI Assistance Challenge Weihua Du, Qiushi Lyu, Jiaming Shan, Zhenting Qi, Hongxin Zhang, Sunli Chen, Andi Peng , Tianmin Shu, Kwonjoon Lee, Behzad Dariush, Chuang Gan. <i>Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks</i> , 2024
ICML 2024	Pragmatic Feature Preferences: Learning Reward-Relevant Preferences from Human Feedback Andi Peng , Yuying Sun, Tianmin Shu, David Abel. <i>International Conference on Machine Learning (ICML)</i> , 2024 <i>RLC Workshop on Reinforcement Learning Beyond Rewards</i> , 2024 Press: JHU News
ICLR 2024	Learning with Language-Guided State Abstractions Andi Peng , Ilia Sucholutsky*, Belinda Z. Li*, Theodore R. Sumers, Thomas L. Griffiths, Jacob Andreas, Julie A. Shah. <i>International Conference on Learning Representations (ICLR)</i> , 2024 <i>RSS Workshop on Social Intelligence in Humans and Robots</i> , 2023 (oral) Press: MIT News
HRI 2024	Preference-Conditioned Language-Guided Abstraction Andi Peng , Andreea Bobu, Belinda Z. Li, Theodore R. Sumers, Ilia Sucholutsky, Thomas L. Griffiths, Julie A. Shah. <i>ACM/IEE International Conference on Human Robot Interaction (HRI)</i> , 2024 <i>ICLR Workshop on LLM Agents</i> , 2024 <i>New England NLP Meeting Series</i> , 2024

HRI 2024	Aligning Human and Robot Representations Andreea Bobu*, Andi Peng *, Pulkit Agrawal, Julie A. Shah, Anca D. Dragan. <i>ACM/IEE International Conference on Human Robot Interaction (HRI)</i> , 2024 <i>ICRA Workshop on Collaborative Robots and Work of the Future</i> , 2022 <i>RSS Workshop on Social Intelligence in Humans and Robots</i> , 2022 <i>NeurIPS Workshop on ML Safety</i> , 2022
NeurIPS 2023	Human-Guided Complexity-Controlled Abstractions Andi Peng *, Mycal Tucker*, Eoin M. Kenny, Noga Zaslavsky, Pulkit Agrawal, Julie A. Shah. <i>Advances in Neural Information Processing Systems (NeurIPS)</i> , 2023
ICML 2023	Diagnosis, Feedback, Adaptation: A Human-in-the-Loop Framework for Test-Time Policy Adaptation Andi Peng , Aviv Netanyahu, Mark K. Ho, Tianmin Shu, Andreea Bobu, Julie A. Shah, Pulkit Agrawal. <i>International Conference on Machine Learning (ICML)</i> , 2023 <i>NeurIPS Workshop on Human in the Loop Learning</i> , 2022 Press: MIT News (front page featured story), EE Times, ASME
AAAI 2022 (oral, top 4%)	Investigations of Performance and Bias in Human-AI Teamwork in Hiring Andi Peng , Besmira Nushi, Kori Inkpen, Emre Kiciman, Ece Kamar. <i>AAAI Conference on Artificial Intelligence (AAAI)</i> , 2022 <i>CHI Workshop on Trust and Reliance in Human-AI Teams</i> , 2022 (oral)
AAAI 2020 (oral, top 3%)	Human-Machine Collaboration for Fast Land-Cover Mapping Caleb Robinson, Anthony Ortiz, Kolya Malkin, Blake Elias, Andi Peng , Dan Morris, Bistra Dilkina, Nebojsa Jovic. <i>AAAI Conference on Artificial Intelligence (AAAI)</i> , 2020 <i>ICLR Workshop on Climate Science and Adaptation</i> , 2020 <i>NeurIPS Workshop on Tackling Climate Change with Machine Learning</i> , 2019
AIES 2020 (spotlight)	The Perils of Objectivity: Towards a Normative Framework for Fair Judicial Decision-Making Andi Peng , Malina Simard-Halm. <i>AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)</i> , 2020
HCOMP 2019	What You See is What You Get? The Impact of Representation Criteria on Human Bias in Hiring Andi Peng , Besmira Nushi, Emre Kiciman, Kori Inkpen, Siddharth Suri, Kori Inkpen, Ece Kamar. <i>AAAI Conference on Human Computation and Crowdsourcing (HCOMP)</i> , 2019
SciTech 2017 (oral)	Conceptual Feasibility Study of the Hyperloop Vehicle for Next-Generation Transport Kenneth Decker, Jeffrey Chin, Andi Peng , Colin Summers, Golda Nguyen, Andrew Oberlander, Gazi Sakib, Nariman Sharifrazi, Christopher Heath, Justin S. Gray, Robert D. Falck. <i>AIAA SciTech Forum and Exposition (SciTech)</i> , 2017 Archived as NASA Technical Report

Workshop Publications

ICLR 2022	Strengthening Subcommunities: Towards Sustainable Growth in AI Research Andi Peng , Jessica Zosa Forde, Yonadav Shavit, Jonathan Frankle. <i>ICLR Workshop on ML Evaluation Standards</i> , 2022
CHI 2020 (oral)	On the Nature of Bias Percolation: Assessing Multiaxial Collaboration in Human-AI Systems Andi Peng , Besmira Nushi, Kori Inkpen, Emre Kiciman, Ece Kamar. <i>CHI Workshop on Human-Centered Approaches to Fair and Responsible AI</i> , 2020

Policy Reports

Apr 2022	Led and evaluated grant on Improving the ML Publishing Process. Schmidt Futures. ICLR 2022 ML Evaluation Standards Workshop.
Apr 2019	Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System The Partnership on AI. Working Group on Fairness, Transparency, and Accountability.

Sep 2018	National Strategic Overview for Quantum Information Science The White House. Office of Science and Technology Policy.
Dec 2016	Nigeria: Tracking and Promoting Good Governance United States Institute of Peace. Through the Yale Jackson School of Global Affairs.

Industry Experience

2021 - 2022	Policy Analyst, Schmidt Futures (Part-time)	New York, NY
2019 - 2020	Applied Scientist II, Microsoft AI & Research	Redmond, WA
2018 - 2019	AI Resident, Microsoft Research	Redmond, WA
Summer 2017	Security Engineering Intern, Facebook eCrime Team <i>Created ML threat modeling to aid federal investigators. Collaborated with law enforcement on counter-terrorism, sex trafficking, and state-sponsored information cases.</i>	Menlo Park, CA
2014 - 2015	Product Manager, IT Central Station	Tel Aviv, Israel

Invited Talks

Jul 2025	far.ai Seminar
Dec 2024	NeurIPS, WiML Panelist
May 2024	ICLR, Representational Alignment Workshop
Feb 2024	UC Berkeley, Jacob Steinhardt Group
Feb 2024	MILA, Robot Learning Seminar
Feb 2024	Constellation, Visiting Researcher Program
Sep 2023	CMU, HRI Reading Group
Jun 2023	Aon, AI Fireside Chat
May 2023	Yale for Humanity: AI, Ethics, and Society: Utilizing Technology for Good
Jun 2022	FAccT, SEDL Workshop
Mar 2022	Yale, Cyber Leadership Forum
Oct 2021	MIT, GW6 Research Summit
Jun 2021	Facebook AI Research, Robotics Seminar
Jun 2020	Microsoft Research, Adaptive Systems and Interaction Group
Mar 2019	Microsoft Research, Diversity, Inclusion and Belonging Day
Feb 2019	Microsoft Research, AI for Good Group
Sep 2018	Microsoft Research, AI Seminar
Aug 2016	NASA, Aeronautics Research Mission Directorate
Aug 2014	Hubei University, School of International Studies

Conference and Workshop Organization

ICRA 2025	Workshop on Safely Leveraging Vision-Language Foundation Models in Robotics	Atlanta, GA
RSS 2025	Program Committee, RSS Pioneers	Los Angeles, CA
AAAI 2025	Workflow Chair	Philadelphia, PA
RSS 2024	Workshop on Social Intelligence in Humans and Robots	Delft, Netherlands
RSS 2024	Workshop on Task Specification for General-Purpose Intelligent Robots	Delft, Netherlands
NeurIPS 2023	Workshop on Goal-Conditioned Reinforcement Learning	New Orleans, LA
ICML 2023	Workshop on Learning from Implicit Human Feedback	Honolulu, HI
CoRL 2022	Workshop on Aligning Robot Representations with Humans	Auckland, New Zealand

Teaching

2022-2024	Yale Jackson School of Global Affairs GLBL 6610: Artificial Intelligence, Emerging Technologies, and National Power	Guest Lecturer (8x)
IAP 2021	MIT Electrical Engineering and Computer Science 6.S090: Deep Learning for Control	Co-Head T.A.
Fall 2017	Yale Computer Science CPSC 100 (CS50): Introduction to Computer Science	T.A.
Spring 2017	CPSC 223: Data Structures and Programming Techniques	T.A.
Spring 2016	CPSC 202: Mathematical Tools for Computer Science	T.A.
Fall 2015, 2016	CPSC 100 (CS50): Introduction to Computer Science <i>First undergraduate head T.A. for the largest engineering course in university history (managed course staff of 62). Had weekly teaching sections professionally filmed and produced for streaming.</i>	Head T.A.
Spring 2015	Yale Astrophysics ASTR 343: Gravity, Astrophysics, and Cosmology	T.A.

Professional Service

	University Service Board of Advisors, Yale Jackson School of Global Affairs
2020 - 2023	
2023 -	Department Service Student Advisory Group, MIT EECS Faculty Search
2023 -	PhD Admissions Student Reviewer, MIT EECS
2021 -	Graduate Visit Days Event Host, MIT EECS
2017 - 2018	Executive Board, Yale Psi Chi Honor Society
2016 - 2018	Student Advisory Board, Yale Jackson Institute for Global Affairs
2017 - 2018	Student Advisory Board, Yale Brady-Johnson Program in Grand Strategy
	Outreach
2022 -	MIT EECS Buddy (underrepresented students in CS)
2016 - 2018	Yale FLOAT (underrepresented students in CS)
	Leadership
2017 - 2018	Peer Liaison, Yale Asian-American Cultural Center <i>Sole upperclassman peer mentor in Berkeley College (one of 14 residential colleges at Yale).</i>
2015 - 2016	Captain, Yale Women's Club Soccer
2013 - 2015	Deputy ED, Teaching Peace Initiative <i>Helped run a national student-run 501(c)(3) nonprofit for teaching peace-curriculum in schools.</i>

Reviewing

	Program Committee AAAI Conference on Artificial Intelligence (AAAI)
2023 -	
2023 -	Reviewing International Conference on Machine Learning (ICML)
2022 -	International Conference on Learning Representations (ICLR)
2024 -	IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
2023 -	Journal of Machine Learning Research (JMLR)
2022 -	Conference on Neural Information Processing Systems (NeurIPS)
2024 -	Annual Meeting of the Cognitive Science Society (CogSci)
2024 -	IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
2024 -	Conference on Robot Learning (CoRL)
2022 -	Robotics: Science and Systems (RSS)
2022 -	IEEE International Conference on Robotics and Automation (ICRA)
2024 -	International Journal of Robotics Research (IJRR)
2021 -	AAAI Conference on Artificial Intelligence (AAAI)

2020, 2024 - ACM Conference on Human Factors in Computing Systems (CHI)
2020 AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)

Research Mentorship

Spring 2025 Sky Wang, Christy Li (MATS Fellows)
Spring 2024 Naomi Kadish, Halli Watson, Alex Wardle-Solano, Joshua Zhu, Juan Otero (Yale Master's Students)
Fall 2023 - Mehul Damani (MIT PhD)
Fall 2023 Yuying Sun (Boston University Undergraduate)
Fall 2023 Weihua Du (Tsinghua Undergraduate)
Summer 2023 Jiaming Shan (MIT-IBM Watson AI Lab Intern)
Fall 2021 Jerry Mao (MIT Undergraduate)

Life Things

2023 - 2025 Boston Marathon (x3)
2022 - Boston Athletic Association (BAA) Running Team
2014 - 2018 Kappa Alpha Theta, Yale
2006 - 2013 Cleveland Institute of Music (theory, chamber, philharmonic student)

Languages (Human)

English	Native
Mandarin	Native
French	Conversational