ANDI PENG

CONTACT

MIT CSAIL (45-711)
32 Vassar Street
Cambridge, MA 02139

andipeng@mit.edu
+1 (440) 715-0384
https://andipeng.com

EDUCATION

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

RESEARCH POSITIONS

Fall 2023	Research Intern, **Boston Dynamics AI Institute**	Cambridge, MA
	Host: Jessica Hodgins	
Summer 2023	Research Intern, **MIT-IBM Watson AI Lab**	Cambridge, MA
	Host: Chuang Gan	
Summer 2021	Research Intern, **Facebook AI Research (FAIR)**	Pittsburgh, PA
	Hosts: Aravind Rajeswaran and Vikash Kumar	
2018 - 2020	AI Resident, **Microsoft Research**	Redmond, WA
	Hosts: Ece Kamar, Besmira Nushi, Emre Kiciman, Kori Inkpen	
2018	Policy Intern, **White House Office of Science and Technology Policy (OSTP)**	Washington, DC
	Research Associate, **National Institute for Standards and Technology (NIST)**	
	Host: Jake Taylor	
2016 - 2018	Undergraduate Researcher, **Yale University**	New Haven, CT
	Advisor: Brian Scassellati	
Summer 2016	MARTI Researcher, **NASA Glenn Research Center**	Cleveland, OH
	Hosts: Justin Gray and Jeffrey Chin	

FELLOWSHIPS, HONORS, AND AWARDS

2023 - 2024	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
	A special distinction that reflects the university's esteem for past and future achievements	
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)*
	Nominated by Senator Sherrod Brown and Congressman Steve LaTourette	

JOURNAL PUBLICATIONS

Under Review **Getting Aligned on Representational Alignment**
Ilia Sucholutsky*, Lukas Muttenthaler*, Adrian Weller, **Andi Peng**, more authors.

TMLR **Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback**
Stephen Casper*, Xander Davies*, ..., **Andi Peng**, more authors.
Transactions on Machine Learning Research (TMLR), 2023

Nature **Make Greenhouse-Gas Accounting Reliable—Build Interoperable Systems**
Amy Luers, Leehi Yona, ..., **Andi Peng**, Christian Stoll, Marian Van Pelt, Ross Virginia, Lucas Joppa.
Nature 607.7920, 2022

CONFERENCE PUBLICATIONS

* equal contribution

ICLR 2024 Learning with Language-Guided State Abstractions
Andi Peng, Ilia Sucholutsky*, Belinda Li*, Theodore Sumers, Thomas Griffiths, Jacob Andreas*, Julie Shah*.
Workshop on Social Intelligence in Humans and Robots, RSS 2023 **(oral)**
Proceedings of the *International Conference on Learning Representations* (ICLR), 2024

HRI 2024 Preference-Conditioned Language-Guided Abstraction
Andi Peng, Andreea Bobu, Belinda Li, Theodore Sumers, Ilia Sucholutsky, Thomas Griffiths, Julie Shah.
Proceedings of the *ACM/IEE International Conference on Human Robot Interaction* (HRI), 2024

HRI 2024 Aligning Human and Robot Representations
Andreea Bobu*, **Andi Peng***, Pulkit Agrawal, Julie Shah, Anca Dragan.
Proceedings of the *ACM/IEE International Conference on Human Robot Interaction* (HRI), 2024
Workshop on Collaborative Robots and Work of the Future, ICRA 2022
Workshop on Social Intelligence in Humans and Robots, RSS 2022
Workshop on ML Safety, NeurIPS 2022

NeurIPS 2023 **Human-Guided Complexity-Controlled Abstractions**
Andi Peng*, Mycal Tucker*, Eoin Kenny, Noga Zaslavsky, Pulkit Agrawal, Julie Shah.
Advances in Neural Information Processing Systems (NeurIPS), 2023

ICML 2023 **Diagnosis, Feedback, Adaptation: A Human-in-the-Loop Framework for Test-Time Policy Adaptation**
Andi Peng, Aviv Netanyahu, Mark Ho, Tianmin Shu, Andreea Bobu, Julie Shah, Pulkit Agrawal.
Proceedings of the *International Conference on Machine Learning* (ICML), 2023
Workshop on Human in the Loop Learning, NeurIPS 2022
Press: **MIT News** (front page featured story), EE Times, ASME

AAAI 2022 **Investigations of Performance and Bias in Human-AI Teamwork in Hiring**
(oral, top 4%) **Andi Peng**, Besmira Nushi, Kori Inkpen, Emre Kiciman, Ece Kamar.
Proceedings of the *AAAI Conference on Artificial Intelligence* (AAAI), 2022
Workshop on Trust and Reliance in Human-AI Teams, CHI 2022 **(oral)**

AAAI 2020 **Human-Machine Collaboration for Fast Land-Cover Mapping**
(oral, top 3%) Caleb Robinson, Anthony Ortiz, Kolya Malkin, Blake Elias, **Andi Peng**, Dan Morris, Bistra Dilkina, Nebojsa Jojic.
Proceedings of the *AAAI Conference on Artificial Intelligence* (AAAI), 2020
Workshop on Climate Science and Adaptation, ICLR 2020
Workshop on Tackling Climate Change with Machine Learning, NeurIPS 2019

AIES 2020 **The Perils of Objectivity: Towards a Normative Framework for Fair Judicial Decision-Making**
(spotlight) **Andi Peng**, Malina Simard-Halm.
Proceedings of the *AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society* (AIES), 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.
Proceedings of the *AAAI Conference on Human Computation and Crowdsourcing* (HCOMP), 2019

SciTech 2017 **Conceptual Feasibility Study of the Hyperloop Vehicle for Next-Generation Transport**
(oral) Kenneth Decker, Jeffrey Chin, **Andi Peng**, Colin Summers, Golda Nguyen, Andrew Oberlander, Gazi

Sakib, Nariman Sharifrazi, Christopher Heath, Justin Gray, Robert Falck.
Proceedings of the *AIAA SciTech Forum and Exposition* (SciTech), 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.
Workshop on ML Evaluation Standards, ICLR 2022

CHI 2020   On the Nature of Bias Percolation: Assessing Multiaxial Collaboration in Human-AI Systems
(oral)   **Andi Peng**, Besmira Nushi, Kori Inkpen, Emre Kiciman, Ece Kamar.
Workshop on Human-Centered Approaches to Fair and Responsible AI, CHI 2020

POLICY WORK

P4.   Led and evaluated grant on Improving the ML Publishing Process.
Schmidt Futures. ICLR 2022 ML Evaluation Standards Workshop.
Apr 2022

P3.   Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System
The Partnership on AI. Working Group on Fairness, Transparency, and Accountability.
Apr 2019

P2.   National Strategic Overview for Quantum Information Science
The White House. Office of Science and Technology Policy.
Sep 2018

P1.   Nigeria: Tracking and Promoting Good Governance
United States Institute of Peace. Through the Yale Jackson School of Global Affairs.
Dec 2016

INDUSTRY EXPERIENCE

2023 -	(Part-time) Special Government Employee (SGE), **Defense Innovation Unit (DIU)**	Washington, DC
2021 - 2022	(Part-time) Policy Analyst, **Schmidt Futures**	New York, NY
2019 - 2020	Applied Scientist II, **Microsoft AI & Research**	Redmond, WA
Summer 2017	Security Engineering Intern, **Facebook eCrime Team**	Menlo Park, CA
	Created ML threat modeling to aid federal investigators. Collaborated with law enforcement on counter-terrorism, sex trafficking, and state-sponsored information cases.	
2014 - 2015	Product Manager, **IT Central Station**	Tel Aviv, Israel

INVITED TALKS

2024	MILA Robot Learning Seminar	Virtual
2023	CMU HRI Reading Group	Virtual
2023	Aon AI Fireside Chat	Chicago, IL
2023	Yale for Humanity: AI, Ethics, and Society: Utilizing Technology for Good	Chicago, IL
2022	FAccT SEDL Workshop	Seoul, South Korea
2022	Yale Cyber Leadership Forum	New Haven, CT
2021	MIT GW6 Research Summit	Cambridge, MA
2021	Facebook AI Research (FAIR) Robotics Seminar	Pittsburgh, PA
2020	Microsoft Research Adaptive Systems and Interaction Group	Virtual
2019	Microsoft AI&R Diversity, Inclusion and Belonging Day	Redmond, WA
2019	Microsoft AI for Good Research Group	Redmond, WA
2018	Microsoft Research AI Seminar	Redmond, WA

| 2016 | NASA Aeronautics Research Mission Directorate | Glenn, OH |
| 2014 | Hubei University School of International Studies | Enshi, China |

WORKSHOP ORGANIZATION

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

Yale Jackson School of Global Affairs
2022-2024   GLBL 6610: Artificial Intelligence, Emerging Technologies, and National Power   Guest Lecturer (8x)

MIT Electrical Engineering and Computer Science
IAP 2021   6.S090: Deep Learning for Control   Co-Head T.A.

Yale Computer Science
Fall 2017	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   Head T.A.
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.

Yale Astrophysics
Spring 2015   ASTR 343: Gravity, Astrophysics, and Cosmology   T.A.

PROFESSIONAL SERVICE

University Service
2020 - 2023   Board of Advisors, Yale Jackson School of Global Affairs

Department Service
2023 -	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
2016 - 2018   Yale FLOAT (Women and Minorities in CS)

Leadership
2017 - 2018   Peer Liaison, Yale Asian-American Cultural Center
Sole upperclassman peer mentor in Berkeley College (one of 14 residential colleges at Yale).
2015 - 2016   Captain, Yale Women's Club Soccer
2013 - 2015   Deputy ED, Teaching Peace Initiative
Helped run a national student-run 501(c)(3) nonprofit for teaching peace-curriculum in schools.

REVIEWING

Program Committee
2023 -   AAAI Conference on Artificial Intelligence (AAAI)

Reviewing
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)
2022 -	Robotics: Science and Systems (RSS)
2022 -	IEEE International Conference on Robotics and Automation (ICRA)
2021 - 2022	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 2023 Mehul Damani (MIT PhD)
Research on learning from human feedback.

Fall 2023 Yuying Sun (BU Undergraduate)
Research on learning from feature preferences.

Fall 2023 Weihua Du (Tsinghua University Undergraduate)
Research on collaborative agents.

Summer-Fall 2023 Jiaming Shan (MIT-IBM Watson AI Lab Undergraduate)
Research on collaborative agents.

Fall 2021 Jerry Mao (MIT Undergraduate)
Research on DARPA Machine Common Sense Project.

LIFE THINGS

2023 - 2024	Boston Marathon (x2)
2022 -	Boston Athletic Association (BAA) Running Team
2017 -	Agora Society, Yale
2014 - 2018	Kappa Alpha Theta, Yale
2006 - 2013	Cleveland Institute of Music (theory, chamber, philharmonic student)

LANGUAGES (HUMAN)

English	Native
Mandarin	Native
French	Conversational