

ANDI PENG

CONTACT

MIT CSAIL (32-386)
32 Vassar Street
Cambridge, MA 02139

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

EDUCATION

| | | |
|----------|---|---------------|
| Ph.D. | Massachusetts Institute of Technology Electrical Engineering and Computer Science Advisors: Jacob Andreas and Julie Shah | Cambridge, MA |
| Feb 2023 | Massachusetts Institute of Technology Electrical Engineering and Computer Science Advisors: Pulkit Agrawal and Julie Shah | Cambridge, MA |
| May 2018 | Yale University , <i>cum laude, with distinction</i> 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 | New Haven, CT |

RESEARCH POSITIONS

| | | |
|-------------|--|----------------|
| 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 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 Intern, White House Office of Science and Technology Policy (OSTP) Research Associate, National Institute for Standards and Technology (NIST) Host: Jake Taylor | Washington, DC |
| 2016 - 2018 | Undergraduate Researcher, Yale University Advisor: Brian Scassellati | New Haven, CT |
| Summer 2016 | MARTI Researcher, NASA Glenn Research Center Hosts: Justin Gray and Jeffrey Chin | Cleveland, OH |

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 (<i>declined</i>) |
| 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 <i>A special distinction that reflects the university's esteem for past and future achievements</i> | \$55,000 |
| 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 <i>Nominated by Senator Sherrod Brown and Congressman Steve LaTourette</i> | (<i>declined</i>) |

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 Summers, 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 Summers, Ilia Sucholutsky, Thomas Griffiths, Julie Shah
Proceedings of the ACM/IEEE 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/IEEE 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](#)
Kenneth Decker, Jeffrey Chin, **Andi Peng**, Colin Summers, Golda Nguyen, Andrew Oberlander, Gazi

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 <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

| | | |
|------|---|--------------------|
| 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

| | | |
|-----------------|---|---------------------|
| 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

| | |
|--|-------------|
| Student Advisory Group, MIT EECS Faculty Search | 2023 |
| Board of Advisors, Yale Jackson School of Global Affairs | 2020 - 2023 |

Department Service

| | |
|--|-------------|
| PhD Admissions Student Reviewer, MIT EECS | 2023 - |
| Graduate Visit Days Event Host, MIT EECS | 2021 - |
| Executive Board, Yale Psi Chi Honor Society | 2017 - 2018 |
| Student Advisory Board, Yale Jackson Institute for Global Affairs | 2016 - 2018 |
| Student Advisory Board, Yale Brady-Johnson Program in Grand Strategy | 2017 - 2018 |

Outreach

| | |
|---|-------------|
| Yale FLOAT (Women and Minorities in CS) | 2016 - 2018 |
|---|-------------|

Leadership

| | |
|--|-------------|
| Peer Liaison, Yale Asian-American Cultural Center | 2017 - 2018 |
| <i>Sole upperclassman peer mentor in Berkeley College (one of 14 residential colleges at Yale).</i> | |
| Captain, Yale Women's Club Soccer | 2015 - 2016 |
| Deputy ED, Teaching Peace Initiative | 2013 - 2015 |
| <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 - 2024 |
|---|-------------|

Reviewing

| | |
|---|------|
| International Conference on Learning Representations (ICLR) | 2024 |
| IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) | 2024 |
| Journal of Machine Learning Research (JMLR) | 2023 |
| Conference on Neural Information Processing Systems (NeurIPS) | 2022 |

| | |
|---|-----------------|
| International Conference on Learning Representations (ICLR) | 2022 |
| Annual Meeting of the Cognitive Science Society (CogSci) | 2024 |
| Robotics: Science and Systems (RSS) | 2022 - 2023 |
| IEEE International Conference on Robotics and Automation (ICRA) | 2022 |
| AAAI Conference on Artificial Intelligence (AAAI) | 2021 - 2022 |
| ACM Conference on Human Factors in Computing Systems (CHI) | 2020-2021, 2024 |
| AAAI/ ACM Conference on Artificial Intelligence, Ethics, and Society (AIES) | 2020 |

RESEARCH MENTORSHIP

| | |
|---------------------|---|
| Spring 2023 | Mehul Damani (MIT PhD) <i>Research on multi-agent communication.</i> |
| Fall 2023 | Yuying Sun (BU Undergraduate) <i>Research on learning from feature preferences.</i> |
| Fall 2023 | Weihua Du (Tsinghua University Undergraduate) <i>Research on collaborative agents.</i> |
| Summer-Fall 2023 | Jiaming Shan (MIT-IBM Watson AI Lab Undergraduate) <i>Research on collaborative agents.</i> |
| Fall 2021 | Jerry Mao (MIT Undergraduate) <i>Research on DARPA Machine Common Sense Project.</i> |

TECHNICAL SKILLS

| | |
|-----------|---|
| Languages | Python, Java, C/C++, R, JavaScript, \LaTeX |
| Software | PyTorch, TensorFlow, Gym, OpenCV, Stata |

OTHER INTERESTS

Running team member of the Boston Athletic Association (BAA). Annual participant at the Boston Marathon and USA Track and Field running circuit. Ex-college soccer player.

LANGUAGES (HUMAN)

| | |
|----------|----------------|
| English | Native |
| Mandarin | Native |
| French | Conversational |