Himabindu Lakkaraju

Contact Information

442 Morgan Hall
15 Harvard Way
Boston, MA 02163
E-mail: hlakkaraju@hbs.edu; hlakkaraju@seas.harvard.edu
Webpage: http://himalakkaraju.github.io

Research Interests

Transparency, Fairness, and Safety in Artificial Intelligence (AI); Applications of AI to Healthcare, Law, and Policy; AI for Decision-Making.

Academic & Professional Experience

Harvard University *01/2020 - Present*
Assistant Professor with appointments in Business School and
Department of Computer Science (Affiliate)

Fiddler AI *06/2021 - Present*
Chief AI Research Fellow

Harvard University *11/2018 - 12/2019*
Postdoctoral Fellow

Stanford University *9/2012 - 9/2018*
Research Assistant

Microsoft Research, Redmond *5/2017 - 6/2017*
Visiting Researcher

Microsoft Research, Redmond *6/2016 - 9/2016*
Research Intern

University of Chicago *6/2014 - 8/2014*
Data Science for Social Good Fellow

IBM Research - India, Bangalore *7/2010 - 7/2012*
Research Engineer

SAP Research, Bangalore *7/2009 - 3/2010*
Visiting Researcher

Adobe Systems Pvt. Ltd., Bangalore *7/2007 - 7/2008*
Software Engineer

Education

Stanford University *9/2012 - 9/2018*
Ph.D. in Computer Science
Thesis: Enabling Machine Learning for High-Stakes Decision-Making

Stanford University *9/2012 - 9/2015*
Master of Science (MS) in Computer Science

Indian Institute of Science (IISc) *8/2008 - 7/2010*
Master of Engineering (MEng) in Computer Science & Automation
Thesis: Exploring Topic Models for Understanding Sentiments Expressed in Customer Reviews

Selected Honors & Achievements

JP Morgan Faculty Research Award	2022
Best Paper Award, **ICML Workshop on Interpretable ML in Healthcare**	2022
Amazon Research Award	2021
National Science Foundation (NSF) Amazon Fairness in AI Grant	2021

Google AI for Social Good Research Award 2021

Best Paper Runner Up, **ICML Workshop on Algorithmic Recourse** 2021

Google Research Award 2020

Co-founded **Trustworthy ML Initiative** with the goal of enabling easy access to resources on trustworthy ML & to build a community of researchers/practitioners 2020

Hoopes Prize for undergraduate thesis mentoring, Harvard University 2020

Named as one of the **35 Innovators Under 35** by MIT Tech Review 2019

Named as one of the **Innovators to Watch** by Vanity Fair 2019

Selected for the prestigious **Cowles Fellowship** by Yale University (declined) 2018

INFORMS Data Mining Best Paper Award 2017

Microsoft Research Dissertation Grant 2017

Named as one of the **Rising Stars in Computer Science** 2016

Outstanding Reviewer Award 2016
International World Wide Web Conference (WWW)

Google Anita Borg Fellowship in recognition of research and leadership 2015

Stanford Graduate Fellowship for exceptional academic performance 2013-17

Eminence and Excellence Award for outstanding contributions to research 2012
IBM Research

Research Division Award recognizing research contributions 2012
IBM Research

Best Paper Award, SIAM International Conference on Data Mining (SDM) 2011
"Exploiting Coherence for the Simultaneous Discovery of Latent Facets and associated Sentiments"

SPOT Award for outstanding product contributions 2009
Adobe Systems Pvt. Ltd.

All India Rank 32 (99.82%ile) 2008
Graduate Aptitude Test in Engineering (GATE)
Entrance examination for IISc & IITs in Computer Science & Engineering

University Rank 10, Bachelor of Engineering, Computer Science 2007
Out of 8000 students from 175 colleges

Selected Grants & Fellowships

As Faculty

JP Morgan Faculty Research Award (US$110,000) – Sole PI	2022 - 2024
D3 Institute at Harvard Grant (US$600,000) – PI	2022 - 2025
NSF-Amazon Fairness in AI (FAI) grant (US$375,000) – co-PI	2021 - 2024
Amazon Faculty Research Award (US$70,000) – Sole PI	2021 - 2024
Google AI for Social Good Research Award (US$10,000) – Sole PI	2021 - 2022
Google Faculty Research Award (US$600,000) – PI	2020 - 2024
National Science Foundation (NSF) RI Small (US$450,000) – Harvard PI	2020 - 2023
Bayer Trust in Science Award (US$100,000) – PI	2020 - 2021

As Student

Microsoft Research Dissertation Grant (US$20,000)	2017
Stanford Graduate Fellowship (tuition + US$41,700 p.a.)	2013 - 2017
Google Anita Borg Scholarship (US$10,000)	2015
Facebook Graduate Fellowship Finalist (US$500)	2013

Indian Institute of Science Graduate Scholarship (tuition + Rs.96,000 p.a.)	2008 - 2010
SAP India Research Grant (Rs.150,000)	2009 - 2010

Publications

Total Citations: **4155**

Book Chapters

[46] Analyzing Human Decisions and Machine Predictions in Bail Decision Making
Jon Kleinberg, **Himabindu Lakkaraju**, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan
(author names are ordered alphabetically)
The Inequality Reader: Contemporary and Foundational Readings in Race, Class, and Gender; Third Edition, 2022 (Forthcoming)

Articles in peer-reviewed journals

[45] Human Decisions and Machine Predictions
Jon Kleinberg, **Himabindu Lakkaraju**, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan
QJE - Quarterly Journal of Economics, 2018
(author names are ordered alphabetically)
Featured in MIT Technology Review, Harvard Business Review, The New York Times, and as Research Spotlight on National Bureau of Economics front page

[44] Extracting Latent Personality Traits from Digital Footprints
Michal Kosinski, Yilun Wang, **Himabindu Lakkaraju**, Jure Leskovec
Psychological Methods - 2016

Articles in peer-reviewed conference proceedings

[43] Data Poisoning Attacks on Off-Policy Evaluation Methods
Elita Lobo, Harvineet Singh, Marek Petrik, Cynthia Rudin, **Himabindu Lakkaraju**
UAI - Conference on Uncertainty in Artificial Intelligence, 2022.

[42] Exploring Counterfactual Explanations Through the Lens of Adversarial Examples: A Theoretical and Empirical Analysis
Martin Pawelczyk, Chirag Agarwal, Shalmali Joshi, Sohini Upadhyay, **Himabindu Lakkaraju**
AISTATS - International Conference on Artificial Intelligence and Statistics, 2022.

[41] Probing GNN Explainers: A Rigorous Theoretical and Empirical Analysis of GNN Explanation Methods
Chirag Agarwal, Marinka Zitnik*, **Himabindu Lakkaraju***
AISTATS - International Conference on Artificial Intelligence and Statistics, 2022.

[40] Fairness via Explanation Quality: Evaluating Disparities in the Quality of Post hoc Explanations
Jessica Dai, Sohini Upadhyay, Ulrich Aivodji, Stephen Bach, **Himabindu Lakkaraju**
AIES - AAAI/ACM Conference on AI, Ethics, and Society, 2022.

[39] Towards Robust Off-Policy Evaluation via Human Inputs
Harvineet Singh, Shalmali Joshi, Finale Doshi-Velez, **Himabindu Lakkaraju**
AIES - AAAI/ACM Conference on AI, Ethics, and Society, 2022.

[38] Towards Robust and Reliable Algorithmic Recourse
Sohini Upadhyay*, Shalmali Joshi*, **Himabindu Lakkaraju**
NeurIPS - Advances in Neural Information Processing Systems (NeurIPS), 2021.
Best Paper Runner Up, ICML Workshop on Algorithmic Recourse, 2021.

[37] Reliable Post hoc Explanations: Modeling Uncertainty in Explainability
Dylan Slack, Sophie Hilgard, Sameer Singh, **Himabindu Lakkaraju**
NeurIPS - Advances in Neural Information Processing Systems, 2021.

[36] Counterfactual Explanations Can Be Manipulated
Dylan Slack, Sophie Hilgard, **Himabindu Lakkaraju**, Sameer Singh
NeurIPS - Advances in Neural Information Processing Systems, 2021.

[35] Learning Models for Algorithmic Recourse
Alexis Ross, **Himabindu Lakkaraju**, Osbert Bastani
NeurIPS - Advances in Neural Information Processing Systems, 2021.

[34] Towards the Unification and Robustness of Perturbation and Gradient Based Explanations
Sushant Agarwal, Shahin Jabbari, Chirag Agarwal*, Sohini Upadhyay*, Steven Wu, **Himabindu Lakkaraju**
ICML - International Conference on Machine Learning, 2021.

[33] Towards a Unified Framework for Fair and Stable Graph Representation Learning
Chirag Agarwal, **Himabindu Lakkaraju***, Marinka Zitnik*
UAI - Conference on Uncertainty in Artificial Intelligence, 2021.

[32] Does Fair Ranking Improve Minority Outcomes? Understanding the Interplay of Human and Algorithmic Biases in Online Hiring
Tom Suhr, Sophie Hilgard, **Himabindu Lakkaraju**
AIES - AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, 2021

[31] Fair influence maximization: A welfare optimization approach
Aida Rahmattalabi, Shahin Jabbari, **Himabindu Lakkaraju**, Phebe Vayanos, Eric Rice, Milind Tambe
AAAI - AAAI International Conference on Artificial Intelligence, 2021

[30] Beyond Individualized Recourse: Interpretable and Interactive Summaries of Actionable Recourses
Kaivalya Rawal, **Himabindu Lakkaraju**
NeurIPS - Advances in Neural Information Processing Systems, 2020

[29] Incorporating Interpretable Output Constraints in Bayesian Neural Networks
Wanqian Yang, Lars Lorch, Moritz Gaule, **Himabindu Lakkaraju**, Finale Doshi-Velez
NeurIPS - Advances in Neural Information Processing Systems, 2020

[28] Robust and Stable Black Box Explanations
Himabindu Lakkaraju, Nino Arsov, Osbert Bastani
ICML - International Conference on Machine Learning, 2020
Invited Talk at INFORMS Annual Meeting, 2020

[27] How do I fool you?: Manipulating User Trust via Misleading Black Box Explanations
Himabindu Lakkaraju, Osbert Bastani
AIES - AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, 2020
Invited Talk at INFORMS Annual Meeting, 2020

[26] Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods
Dylan Slack, Sophie Hilgard, Emily Jia, Sameer Singh, **Himabindu Lakkaraju**
AIES - AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, 2020
Featured in Harvard Business Review and deeplearning.ai
Best Paper (Non-Archival) at AAAI Workshop on Safe AI, 2020
Invited Talk at INFORMS Annual Meeting, 2020

[25] Faithful and Customizable Explanations of Black Box Models
Himabindu Lakkaraju, Ece Kamar, Rich Caruana, Jure Leskovec
AIES - AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, 2019
Invited Talk at INFORMS Annual Meeting, 2017

[24] The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables
Himabindu Lakkaraju, Jon Kleinberg, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan
KDD - ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2017

[23] Learning Cost-Effective and Interpretable Treatment Regimes
Himabindu Lakkaraju, Cynthia Rudin
AISTATS - International Conference on Artificial Intelligence and Statistics, 2017

INFORMS Data Mining Best Paper Award, 2017
Invited Talk at INFORMS Annual Meeting, 2017

[22] Identifying Unknown-Unknowns in the Open World: Representations and Policies for Guided Exploration
Himabindu Lakkaraju, Ece Kamar, Rich Caruana, Eric Horvitz
AAAI - AAAI International Conference on Artificial Intelligence, 2017
Featured in Bloomberg Technology

[21] Confusions over Time: An Interpretable Bayesian Model for Characterizing Trends in Decision Making
Himabindu Lakkaraju, Jure Leskovec
NIPS - Advances in Neural Information Processing Systems, 2016

[20] Interpretable Decision Sets: A Joint Framework for Description and Prediction
Himabindu Lakkaraju, Stephen Bach, Jure Leskovec
KDD - ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016
Invited Talk at INFORMS Annual Meeting 2016

[19] A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes
Himabindu Lakkaraju, Everaldo Aguiar, Carl Shan, David Miller, Nasir Bhanpuri, Rayid Ghani, Kecia Addison
KDD - ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015

[18] A Bayesian Framework for Modeling Human Evaluations
Himabindu Lakkaraju, Jure Leskovec, Jon Kleinberg, Sendhil Mullainathan
SDM - SIAM International Conference on Data Mining, 2015

[17] Who, When, and Why: A Machine Learning Approach to Prioritizing Students at Risk of not Graduating High School on Time
Everaldo Aguiar, **Himabindu Lakkaraju**, Nasir Bhanpuri, David Miller, Ben Yuhas, Kecia Addison, Shihching Liu, Marilyn Powell and Rayid Ghani
LAK - Learning Analytics and Knowledge Conference, 2015

[16] What's in a name ? Understanding the Interplay between Titles, Content, and Communities in Social Media
Himabindu Lakkaraju, Julian McAuley, Jure Leskovec
ICWSM - International AAAI Conference on Weblogs and Social Media, 2013
Featured in Time, Forbes, Phys.Org, Business Insider

[15] Dynamic Multi-Relational Chinese Restaurant Process for Analyzing Influences on Users in Social Media
Himabindu Lakkaraju, Indrajit Bhattacharya, Chiranjib Bhattacharyya
ICDM - IEEE International Conference on Data Mining, 2012

[14] Attention prediction on social media brand pages
Himabindu Lakkaraju, Jitendra Ajmera
CIKM - ACM Conference on Information and Knowledge Management, 2011

[13] Exploiting Coherence for the Simultaneous Discovery of Latent Facets and associated Sentiments
Himabindu Lakkaraju, Chiranjib Bhattacharyya, Indrajit Bhattacharya, Srujana Merugu
SDM - SIAM International Conference on Data Mining, 2011
Best Paper Award

[12] TEM: A novel perspective to modeling content on microblogs
Himabindu Lakkaraju, Hyung-Il-Ahn
WWW - International World Wide Web Conference, 2011

[11] Smart news feeds for social networks using scalable joint latent factor models
Himabindu Lakkaraju, Angshu Rai, Srujana Merugu
WWW - International World Wide Web Conference, 2011

Preprints

[10] Which Explanation Should I Choose? A Function Approximation Perspective to Characterizing Post hoc Explanations [PDF]
Tessa Han, Suraj Srinivas, **Himabindu Lakkaraju**
Best Paper Award, ICML Workshop on Interpretable ML in Healthcare

[9] The Disagreement Problem in Explainable Machine Learning:
A Practitioner's Perspective [PDF]
Satyapriya Krishna, Tessa Han, Alex Gu, Javin Pombra, Shahin Jabbari, Steven Wu, **Himabindu Lakkaraju**
Featured in Fortune Magazine

[8] OpenXAI: Towards a Transparent Evaluation of Model Explanations [PDF]
Chirag Agarwal, Satyapriya Krishna, Eshika Saxena, Nari Johnson, Martin Pawelczyk, Isha Puri, Marinka Zitnik, **Himabindu Lakkaraju**

[7] Rethinking Explainability as a Dialogue: A Practitioner's Perspective [PDF]
Himabindu Lakkaraju, Dylan Slack, Yuxin Chen, Chenhao Tan, Sameer Singh

[6] Let Users Decide: Navigating the Trade-Offs Between Costs and Robustness in Algorithmic Recourse [PDF]
Martin Pawelczyk, Teresa Datta, Johannes van den Heuvel, Gjergji Kasneci, **Himabindu Lakkaraju**

[5] TalkToModel: Understanding Machine Learning Models With Open Ended Dialogues PDF
Dylan Slack, Satyapriya Krishna, **Himabindu Lakkaraju***, Sameer Singh*

[4] When Does Uncertainty Matter?: Understanding the Impact of Predictive Uncertainty in ML Assisted Decision Making [PDF]
Sean McGrath, Parth Mehta, Alexandra Zytek, Isaac Lage, **Himabindu Lakkaraju**
Featured in VentureBeat

[3] Flatten the Curve: Efficiently Training Low-Curvature Neural Networks [PDF]
Suraj Srinivas, Kyle Matoba, **Himabindu Lakkaraju**, Francois Fleuret

Patents

[2] Extraction and Grouping of Feature Words
Chiranjib Bhattacharyya, **Himabindu Lakkaraju**, Sunil Aravindam, Kaushik Nath
US8484228 B2

[1] Enhancing knowledge bases using rich social media
Jitendra Ajmera, Shantanu Godbole, **Himabindu Lakkaraju**, Ashish Verma, Ben Roden
US20130224714 A1

Advising & Mentoring

Current Advisees:

Jiaqi Ma, Postdoctoral Fellow, Harvard University	2022 - Present
Suraj Srinivas, Postdoctoral Fellow, Harvard University	2021 - Present
Abhimanyu Dubey, Visiting Postdoctoral Fellow, Harvard University	2022 - Present
Satyapriya Krishna, PhD Student, Harvard University	2021 - Present
Tessa Han, PhD Student, Harvard University	2020 - Present
Dylan Slack, PhD Student, UC Irvine	2019 - Present
Martin Pawelczyk, PhD Student, University of Tubingen	2021 - Present
Tom Suhr, Masters Student, University of Tubingen	2020 - Present
Isha Puri, Undergrad, Harvard University	2022 - Present
Eshika Saxena, Undergrad, Harvard University	2021 - Present

Current Research Interns:

Umang Bhatt, PhD Student, University of Cambridge	2022
Anna Meyer, PhD Student, University of Wisconsin Madison	2022

Ruijiang Gao, PhD Student, Universtiy of Texas at Austin	2022
Vishwali Mhasawade, PhD Student, New York University	2022

Past Advisees, Visitors, and Interns:

Chirag Agarwal, Postdoctoral Fellow, Harvard University	2020 - 2022
Shahin Jabbari, Postdoctoral Fellow, Harvard University	2019 - 2021
Sophie Hilgard, PhD Student, Harvard University	2019 - 2021
Elita Lobo, PhD Student, University of Massachussetts, Amherst	2020 - 2021
Harvineet Singh, PhD Student, New York University	2020 - 2021
Kaivalya Rawal, MS Student, Harvard University	2019 - 2021
Aditya Karan, MS Student, Harvard University	2019 - 2020
Javin Pombra, Undergrad, Harvard University	2021 - 2022
Ethan Kim, Undergrad, Harvard University	2021
Alexis Ross, Undergrad, Harvard University	2019 - 2021
Jorma Gorns, Undergrad, Harvard University	2019 - 2020
Emily Jia, Undergrad, Harvard University	2019 - 2020
Nino Arsov, Visiting Researcher, Stanford University	2016, 2019 - 2020
Rishabh Bhargava, MS Student, Stanford University	2015
Yilun Wang, MS Student, Stanford University	2014 - 2015

Teaching Experience

Instructor, Interpretability and Explainability in ML Fall 2019 &
Harvard CS & Harvard Business School Spring 2021
First ever course on this emerging topic

Instructor, Technology and Operations Management Fall 2020 &
Harvard Business School Fall 2021

Instructor, Introduction to ML for Social Scientists, Harvard Business School Spring 2020

Instructor, Explainable and Accurate AI for High-Stakes Decision Making
Harvard Business Analytics Program (HBAP) 2020 - 2022

Guest Lecture, Introduction to Data Science, Stanford Law School Spring 2016

Co-instructor, Probability with Mathemagics, Spring 2016
Stanford: Splash Initiative for High School Students

Teaching Assistant, Stanford: Mining Massive Data Sets (CS 246) Winter 2016

Guest Lecture, Algorithms for Submodular Optimization Winter 2016
Stanford: Mining Massive Data Sets (CS 246)

Co-instructor, Introduction to Python Programming Spring 2015
Stanford: Girls Teaching Girls to Code (GTGTC) for High School Students

Mathematics and Science Tutor Winter 2015
DreamCatchers Nonprofit Organization, Palo Alto

Head Teaching Assistant, Autumn 2014
Stanford: Social & Information Network Analysis (CS 224W)

Head Teaching Assistant, Autumn 2010
Indian Institute of Science: Machine Learning

Tutorials

Model Monitoring in Practice: Lessons Learned and Open Challenges FAccT 2022

Explaining Machine Learning Predictions:
State-of-the-art, Challenges, and Opportunities AAAI 2021

Explainable ML in the Wild: When Not to Trust Your Explanations FAccT 2021

Explainable ML: Understanding the Limits and Pushing the Boundaries CHIL 2021
Invited Tutorial

	Explaining Machine Learning Predictions: State-of-the-art, Challenges, and Opportunities	NeurIPS 2020
Invited Talks & Panel Discussions	Amazon Distinguished Lecture	2022
	Stanford Center for AI Safety Workshop on Explainable AI	2022
	CVPR Workshop on Explainable AI for Computer Vision	2022
	Stanford Human-Centered Artificial Intelligence (HAI) Conference	2022
	University of Southern California	2022
	Keynote at WWW Workshop on Explainable AI in Health	2022
	ICLR Workshop on Privacy, Accountability, Interpretability, Robustness, Reasoning on Structured Data	2022
	INFORMS Annual Meeting	2022
	Keynote at ACM CIKM Conference	2021
	NIST AI Risk Management Framework Workshop	2021
	Pinterest Distinguished Lecture	2021
	NeurIPS Workshop on Algorithmic Fairness through the Lens of Causality and Robustness	2021
	NeurIPS Workshop on Explainable AI Approaches for Debugging and Diagnosis	2021
	NeurIPS Workshop on Human and Machine Decisions	2021
	Keynote at ICML Workshop on Interpretable ML in Healthcare	2021
	Keynote at KDD Workshop on ML in finance	2021
	AI for Good Summit organized by International Telecommunications Union & the United Nations	2021
	Keynote at CVPR Workshop on Responsible Computer Vision	2021
	Keynote at ICLR Workshop on Responsible AI	2021
	Keynote at ASPLOS Workshop on Systems Architecture for Robust, Safe, and Resilient Software	2021
	Keynote at MLSys Workshop on Personalized Recommender Systems & Algorithms	2021
	University of Cambridge	2021
	Neurosym Webinar Series, Jointly Organized by UPenn, MIT, Caltech, and Stanford	2021
	Voices of Data Science, UMass Amherst	2021
	Max Planck Symposium on Computing and Society	2021
	Machine Learning Department and Institute of Software Research at Carnegie Mellon University	2020
	Keynote at CVPR Workshop on Fair, Data-Efficient and Trusted Computer Vision	2020
	Keynote at MICCAI Workshop on Interpretability in Medical Imaging	2020
	3 Invited Talks at INFORMS Annual Meeting	2020
	ETH - Center for Law and Economics, Zurich	2020
	University of Michigan, Ann Arbor	2019
	Harvard CRCS Seminar, Cambridge	2019
	INFORMS Annual Meeting, Seattle	2019
	AI World Conference & Expo, Cambridge	2019
	EmTech MIT Conference, Cambridge	2019
	Google DeepMind Annual Summit, Cambridge	2019
	Women in Machine Learning Workshop, Boston	2019
	ICLR Workshop on Safe Machine Learning, New Orleans	2019
	Harvard Data Science Conference, Cambridge	2018
	South Park Commons, San Francisco	2018
	Microsoft Research, Redmond	2018
	Computer Science Department at UCSD, San Diego	2018
	Computer Science Department at University of Michigan, Ann Arbor	2018
	Computer Science Department at Brown University, Providence	2018
	Computer Science Department at UIUC, Urbana Champaign	2018
	Computer Science Department at USC, Los Angeles	2018
	Machine Learning and Computer Science Departments at Carnegie Mellon University, Pittsburgh	2018
	Computer Science Deparment at UCLA, Los Angeles	2018
	Computer Science Deparment at UCI, Irvine	2018

Computer Science Deparment at Duke University, Durham	2018
Computer Science Department at University of Maryland, College Park	2018
NYU Stern School of Business, New York	2018
Operations Research and Information Engineering Department at Cornell University, Ithaca	2018
Industrial Engineering and Operations Research Department at Columbia University, New York	2018
College of Computing at Georgia Tech, Atlanta	2018
Computer Science Department at Harvard University, Cambridge	2018
Computer Science Department at Yale University, New Haven	2018
MIT Sloan School of Management, Cambridge	2018
Harvard Business School, Boston	2018
Operations Research and Financial Engineering Department at Princeton University, Princeton	2018
UC Berkeley School of Public Health, San Francisco	2018
Microsoft Research, Redmond, USA	2017
IBM Thomas J. Watson Research Center, New York	2017
Machine Learning Seminar at Duke University, Durham	2017
INFORMS Annual Meeting, Houston	2017
Keynote at ICML Workshop on Automatic Machine Learning, Sydney, Australia	2017
Stanford Biomedical Data Science Lecture Series, Palo Alto	2017
Stanford Symbolic Systems Coffee Chat Series, Palo Alto	2017
Stanford Data Science Retreat, Palo Alto	2017
Workshop on Demystifying Artificial Intelligence, San Francisco	2017
Disruptive Innovation in Law Conference, Sydney, Australia	2017
Rising Stars Workshop, Pittsburgh	2016
Robert Bosch Research, Palo Alto	2016
INFORMS Annual Meeting, Nashville	2016
Stanford Data Science Retreat, Palo Alto	2016
Future Law: Watson and Beyond (Panel Discussion), Stanford Law School	2016
CodeX Center, Stanford Law School, Palo Alto	2016
KDD Workshop on Data Science for Social Good, New York	2014
University of Chicago Computation Institute, Chicago	2014
Stanford HCI Retreat, San Francisco	2013
Yahoo IR Summer School, Bangalore, India	2011
Indian Institute of Science Talk Series, Bangalore, India	2011
Grace Hopper India Chapter, Bangalore, India	2011

Community Service

Co-Founder & Organizer: Trustworthy ML Initiative
We launched this initiative to enable easy access to resources on trustworthy ML, to showcase and promote the work of researchers from underrepresented groups, and to build a community of researchers and practitioners working on the topic.

Co-Chair:

KDD Trustworthy AI Day	2022
ICML Workshop on New Frontiers in Adversarial Machine Learning	2022
KDD Deep Learning Day	2021
ICML Workshop on Algorithmic Recourse	2021
ELLIS Human-Centric Machine Learning Workshop	2021
Session on Trustworthy Machine Learning at INFORMS	2020
Session on Fairness in Machine Learning at INFORMS	2019
Workshop on Debugging Machine Learning Models at International Conference on Learning Representations (ICLR)	2019
Workshop for spreading awareness about STEM fields among middle school girls	2016
Stanford's Girls Teaching Girls To Code (GTGTC)	2015
Women in Data Science for Social Good Group, UChicago	2014
Grace Hopper India Conference	2011

Area Chair:

NeurIPS - *Advances in Neural Information Processing Systems*	2019 - 2022
ICLR - *International Conference on Learning Representations*	2020 - 2022
AISTATS - *International Conference on Artificial Intelligence and Statistics*	2021 - 2022
ICML - *International Conference on Machine Learning*	2019 - 2021

Program Committee:

AISTATS - *International Conference on Artificial Intelligence and Statistics*	2019 - 2020
AAAI - *AAAI International Conference on Artificial Intelligence*	2019
ICML - *International Conference on Machine Learning*	2018
ICLR - *International Conference on Learning Representations*	2018 - 2019
IJCAI - *International Joint Conference on Artificial Intelligence*	2018 - 2019
WWW - *International World Wide Web Conference*	2017 - 2018
NIPS - *Advances in Neural Information Processing Systems*	2016 - 2017
KDD - *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*	2015 - 2017
CIKM - *ACM Conference on Information and Knowledge Management*	2011, 2017
SDM - *SIAM International Conference on Data Mining*	2015
UAI - *Conference on Uncertainty in Artificial Intelligence*	2011
AAAI - *AAAI conference on Artificial Intelligence*	2011

Journal Reviewer:

OR - *Operations Research*	2021
TWEB - *ACM Transactions on the Web*	2017
PLOS ONE - *Public Library of Science ONE*	2017
TKDD - *ACM Transactions on Knowledge Discovery from Data*	2016
TKDE - *IEEE Transactions on Knowledge and Data Engineering*	2015

Other:

Mentor, Stanford Science Penpals	2017
Member, Ph.D. Student Selection Committee, Stanford Computer Science	2016
Mentor and Sponsor, Children International	2013 - Present
Member, Stanford AI Women Group	2014 - Present

Selected Media Coverage

Fortune: Explainable AI & The Disagreement Problem
Harvard Business Review: The AI transparency paradox
MIT Technology Review: How to upgrade judges with machine learning
Harvard Business Review: Solving social problems with machine learning
The New York Times: Even Imperfect Algorithms Can Improve the Criminal Justice System
VentureBeat: Confidence, uncertainty, and trust in AI affect how humans make decisions
Wired: This Agency Wants to Figure Out Exactly How Much You Trust AI
Bloomberg Technology: Researchers combat gender and racial bias in AI
Forbes: How to craft the perfect Reddit posting
Time: How to succeed on Reddit
Business Insider: How to execute the perfect Reddit submission
Phys.org: Stanford Trio explore success formula for Reddit posts
International Business Times: The secret to what makes something go viral
New Scientist: Things that make a meme explode
The Verge: The math behind successful Reddit submissions
ACM TechNews: Stanford trio explore success formula for Reddit posts
Gizmodo: This equation can tell you how successful a reddit post can be
GigaOm: How to maximize your reddit upvotes, by the numbers