

Himabindu Lakkaraju

Contact Information	Morgan Hall T25 Harvard Business School Soldiers Field Road Boston, MA 02163 <i>E-mail:</i> hlakkaraju@hbs.edu; hlakkaraju@seas.harvard.edu <i>Webpage:</i> http://himalakkaraju.github.io	
Research Interests	Transparency, Fairness, and Safety in Artificial Intelligence (AI); Applications of AI to Law, Healthcare, Public Policy, and Education; AI for Decision-Making.	
Academic & Professional Experience	Harvard University <i>Assistant Professor</i> with appointments in Business School and Department of Computer Science (Affiliate)	01/2020 -
	Harvard University <i>Postdoctoral Fellow</i> at appointments in Business School	11/2018 - 12/2019
	Stanford University <i>Research Assistant</i> in the Department of Computer Science	9/2012 - 9/2018
	Microsoft Research , Redmond <i>Visiting Researcher</i>	5/2017 - 6/2017
	Microsoft Research , Redmond <i>Research Intern</i>	6/2016 - 9/2016
	University of Chicago <i>Data Science for Social Good Fellow</i>	6/2014 - 8/2014
	IBM Research - India , Bangalore <i>Technical Staff Member</i>	7/2010 - 7/2012
	SAP Research , Bangalore <i>Visiting Researcher</i>	7/2009 - 3/2010
	Adobe Systems Pvt. Ltd. , Bangalore <i>Software Engineer</i>	7/2007 - 7/2008
Education	Stanford University Ph.D. in Computer Science Thesis: Enabling Machine Learning for High-Stakes Decision-Making Advisor: Prof. Jure Leskovec Thesis Committee: Prof. Emma Brunskill, Dr. Eric Horvitz, Prof. Jon Kleinberg, Prof. Percy Liang, Prof. Cynthia Rudin	9/2012 - 9/2018
	Stanford University Master of Science (MS) in Computer Science Advisor: Prof. Jure Leskovec	9/2012 - 9/2015
	Indian Institute of Science (IISc) Master of Engineering (MEng) in Computer Science & Automation Thesis: Exploring Topic Models for Understanding Sentiments Expressed in Customer Reviews Advisor: Prof. Chiranjib Bhattacharyya	8/2008 - 7/2010

Publications

Articles in peer-reviewed journals

- [34] 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
- [33] Extracting Latent Personality Traits from Digital Footprints
Michal Kosinski, Yilun Wang, **Himabindu Lakkaraju**, Jure Leskovec
Psychological Methods - 2016

Articles in peer-reviewed conference proceedings

- [32] Robust and Stable Black Box Explanations
Himabindu Lakkaraju, Nino Arsov, Osbert Bastani
ICML - International Conference on Machine Learning, 2020
- [31] 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
- [30] 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
- [29] 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
- [28] 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
- [27] 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 - Finalist, 2017
Invited Talk at INFORMS Annual Meeting, 2017
- [26] 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
- [25] 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
- [24] 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
- [23] 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

- [22] A Bayesian Framework for Modeling Human Evaluations
Himabindu Lakkaraju, Jure Leskovec, Jon Kleinberg, Sendhil Mullainathan
SDM - SIAM International Conference on Data Mining, 2015
- [21] 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
- [20] 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
- [19] 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
- [18] Attention prediction on social media brand pages
Himabindu Lakkaraju, Jitendra Ajmera
CIKM - ACM Conference on Information and Knowledge Management, 2011
- [17] 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

Posters in peer-reviewed conference proceedings

- [16] TEM: A novel perspective to modeling content on microblogs
Himabindu Lakkaraju, Hyung-Il-Ahn
WWW - International World Wide Web Conference, 2011
- [15] 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

Articles in peer-reviewed workshop proceedings

- [14] Interpretable and Explorable Approximations of Black Box Models
Himabindu Lakkaraju, Ece Kamar, Jure Leskovec, Rich Caruana
FAT ML - KDD Workshop on Fairness, Transparency, and Accountability in Machine Learning, 2017
- [13] Learning Interpretable Treatment Regimes
Himabindu Lakkaraju, Cynthia Rudin
NIPS Workshop on Interpretable Machine Learning for Complex Systems, 2016
- [12] Cost-Effective Regimes for Treatment Recommendations
Himabindu Lakkaraju, Cynthia Rudin
NIPS Workshop on Machine Learning for Health, 2016
- [11] Learning Cost-Effective Treatment Regimes for Judicial Bail Decisions
Himabindu Lakkaraju, Cynthia Rudin
NIPS Workshop on Machine Learning and Law, 2016

- [10] Discovering Blind Spots of Predictive Models
Himabindu Lakkaraju, Ece Kamar, Rich Caruana, Eric Horvitz
NIPS workshop on Reliable Machine Learning, 2016
- [9] Using Big Data to Improve Social Policy
Jon Kleinberg, **Himabindu Lakkaraju**, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan
NBER Economics of Crime Working Group, 2014
- [8] Aspect Specific Sentiment Analysis using Hierarchical Deep Learning
Himabindu Lakkaraju, Richard Socher, Chris Manning
NIPS Workshop on Deep Learning and Representation Learning, 2014
- [7] Unified Modeling of User Activities on Social Networking Sites
Himabindu Lakkaraju, Angshu Rai
NIPS Workshop on Computational Social Science and the Wisdom of Crowds, 2011
- [6] A Non-parametric Theme Event Topic Model for Characterizing Microblogs
Himabindu Lakkaraju, Hyung-Il Ahn
NIPS Workshop on Computational Social Science and the Wisdom of Crowds, 2011

Preprints

- [5] Interpretable Explanations for Diverse Clustering
Himabindu Lakkaraju, Nino Arsov, Marinka Zitnik, Jure Leskovec
In preparation
- [4] Rethinking Instrumental Variable Models for Robust Causal Inference
Himabindu Lakkaraju, Cynthia Rudin, Tyler McCormick
In preparation
- [3] Diverse Clustering for Biological Data
Nino Arsov, **Himabindu Lakkaraju**, Marinka Zitnik, Jure Leskovec
In preparation

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](#)

Selected Honors & Awards	Selected 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	2018
	INFORMS Data Mining Best Paper Award - Finalist "Learning Cost-Effective and Interpretable Treatment Regimes"	2017
	Outstanding Reviewer Award International World Wide Web Conference (WWW)	2017
	Named as one of the Rising Stars in EECS	2016
	One of the 25 Ph.D. students invited to Women in Research Lean In with Sheryl Sandberg at Facebook HQ	2016
	Eminence and Excellence Award for outstanding contributions to research IBM Research	2012
	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"

Masters Thesis awarded **Highest Grade** at Indian Institute of Science 2010

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

**Grants
& Fellowships**

Google Faculty Research Award (US\$600,000) 2020
Microsoft Research Dissertation Grant (US\$20,000) 2017
Women in Machine Learning (WiML) Travel Grant for NIPS (US\$850) 2017
ICML Travel Grant (US\$1800) 2017
KDD Travel Grant (US\$1000 p.a.) 2014 - 2017
Stanford Graduate Fellowship (tuition + US\$41,700 p.a.) 2013 - 2017
NIPS Travel Grant (US\$1000) 2016
Google Anita Borg Scholarship (US\$10,000) 2015
Facebook Graduate Fellowship Finalist (US\$500) 2013
Indian Institute of Science Graduate Scholarship 2008 - 2010
(tuition + Rs.96,000 p.a.)
SAP India Research Grant (Rs.150,000) 2009 - 2010
Undergraduate Merit scholarship (Rs.3000 p.a.) 2004 - 2007

**Advising
& Mentoring**

Shahin Jabbari, Postdoctoral Fellow, Harvard University 2019 - Present
Haipeng Cheng, Postdoctoral Fellow, Harvard University 2020 - Present
Sophie Hilgard, PhD Student, Harvard University 2019 - 2020
Maya Balakrishnan, PhD Student, Harvard University 2020 - Present
Dylan Slack, PhD Student, UC Irvine 2019 - 2020
Aida Rahmtalabi, PhD Student, USC 2019 - 2020
Kaivalya Rawal, MS Student, Harvard University 2019 - Present
Aditya Karan, MS Student, Harvard University 2019 - Present
Jorma Gorns, Undergrad, Harvard University 2019 - Present
Alexis Ross, Undergrad, Harvard University 2019 - Present
Emily Jia, Undergrad, Harvard University 2019 - 2020
Wanqian Yang, 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
Mrinal Kanti Das, Ph.D. Student, Indian Institute of Science 2011
Hemant Purohit, Ph.D. Student, Wright State University 2011

**Teaching
Experience**

Instructor, [Introduction to ML for Social Scientists](#), Harvard Business School Spring 2020

Instructor, [Interpretability and Explainability in ML](#), Harvard CS Fall 2019
First course ever on this topic

Guest Lecture, Introduction to Data Science, Stanford Law School Spring 2016
Co-instructor, Probability with Mathemagics, Stanford: Splash Initiative for High School Students Spring 2016

Teaching Assistant, Stanford: Mining Massive Data Sets (CS 246)	Winter 2016
Guest Lecture, Algorithms for Submodular Optimization Stanford: Mining Massive Data Sets (CS 246)	Winter 2016
Co-instructor, Introduction to Python Programming Stanford: Girls Teaching Girls to Code (GTGTC) for High School Students	Spring 2015
Mathematics and Science Tutor DreamCatchers Nonprofit Organization, Palo Alto	Winter 2015
Head Teaching Assistant, Stanford: Social & Information Network Analysis (CS 224W)	Autumn 2014
Head Teaching Assistant, Indian Institute of Science: Machine Learning	Autumn 2010
Co-instructor, Visvesvaraya Technological University: Object Oriented Programming	Autumn 2007
Co-instructor, Visvesvaraya Technological University: Introduction to Databases	Winter 2007
Instructor for English and Mathematics (Grades 8 - 10) UNICEF's Teach India Initiative	2008 - 2010

Invited Talks & Panel Discussions	Keynote at CVPR Workshop on Fair, Data-Efficient and Trusted Computer Vision	2020
	Keynote at MICCAI Workshop on Interpretability in Medical Imaging	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 Department at UCLA, Los Angeles	2018
	Computer Science Department at UCI, Irvine	2018
	Computer Science Department 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, Cambridge	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 **Organizer:**

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:

ICML - <i>International Conference on Machine Learning</i>	2019 - 2020
NeurIPS - <i>Advances in Neural Information Processing Systems</i>	2019 - 2020
ICLR - <i>International Conference on Learning Representations</i>	2020

Program Committee:

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

Journal Reviewer:

TWEB - <i>ACM Transactions on the Web</i>	2017
PLOS ONE - <i>Public Library of Science ONE</i>	2017
EJOR - <i>European Journal of Operational Research</i>	2017
TKDD - <i>ACM Transactions on Knowledge Discovery from Data</i>	2016
TKDE - <i>IEEE Transactions on Knowledge and Data Engineering</i>	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

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](#)
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](#)