

Umang Bhatt

CONTACT INFORMATION	usb20@cam.ac.uk Website	Citizenship: USA UK Global Talent Visa
ACADEMIC POSITIONS	University of Cambridge Assistant Professor, Cambridge Institute of Technology and Humanity Fellow in Computer Science, King's College New York University Assistant Professor & Faculty Fellow, Center for Data Science The Alan Turing Institute Senior Research Associate, Safe and Ethical AI Team Research Associate, Safe and Ethical AI Team Harvard University Research Fellow, Center for Research on Computation and Society	2025 – Present 2025 – Present 2023 – 2025 2024 – 2025 2023 – 2024 2022 – 2023
EDUCATION	University of Cambridge Ph.D. in Engineering <i>Passed with No Corrections</i> <i>Thesis:</i> Trustworthy Machine Learning: From Algorithmic Transparency to Decision Support <i>Committee:</i> Adrian Weller MBE (Advisor), Neil Lawrence, Eric Horvitz <i>Affiliations:</i> Machine Learning Group, Computation and Biological Learning Lab Carnegie Mellon University, Pittsburgh, PA M.S. in Electrical and Computer Engineering B.S. in Electrical and Computer Engineering	2019 – 2023 2017 – 2019 2015 – 2019
SELECT FELLOWSHIPS AND AWARDS	Center for Democracy & Technology Fellowship J.P. Morgan AI PhD Fellowship (<i>Declined</i>) 🔗 The Alan Turing Institute Enrichment Studentship 🔗 Mozilla Fellowship 🔗 Partnership on AI Research Fellowship Leverhulme Center for the Future of Intelligence PhD Scholarship Fully funded by Google DeepMind and the Leverhulme Trust Lovett Family Endowed Scholarship , The Andrew Carnegie Society NSF I-Corps Site Award for research commercialization H. F. McCullough Memorial Scholarship , CMU	2024 – 2026 2022 – 2023 2021 – 2022 2020 – 2021 2019 – 2020 2019 – 2023 2019 2017 2016
SELECT PUBLICATIONS	A more exhaustive list of my publications can be found on Google Scholar . [1] Documenting Deployment with Fabric: A Repository of Real-World AI Governance <i>AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES) 2025</i> Mackenzie Jorgensen, Kendall Brogle, Katherine Collins, Lujain Ibrahim, Arina Shah, Petra Ivanovic, Noah Broestl, Gabriel Piles, Paul Dongha, Hatim Abdulhussein, Adrian Weller, Jillian Powers, Umang Bhatt [2] Learning Personalized Decision Support Policies <i>AAAI Conference on Artificial Intelligence (AAAI) 2025 (Oral Presentation)</i> Umang Bhatt* , Valerie Chen*, Katherine Collins, Parameswaran Kamalaruban, Emma Kallina, Adrian Weller, Ameet Talwalkar [3] Building Machines that Learn and Think <i>with</i> People <i>Nature Human Behavior</i> , 2024. Katherine Collins*, Ilia Sucholutsky*, Umang Bhatt* , Kartik Chandra*, Lionel Wong*, Mina Lee, Cedegao E. Zhang, Tan Zhi-Xuan, Mark Ho, Vikash Mansinghka, Adrian Weller, Joshua Tenenbaum, Thomas Griffiths	

- [4] **When Should Algorithms Resign? A Proposal for AI Governance**
IEEE Computer, 2024
Umang Bhatt*, Holli Sargeant*
- [5] **Evaluating Language Models for Mathematics through Interactions**
Proceedings of the National Academy of Sciences (PNAS), 2024.
 Katherine Collins, Albert Jiang, Simon Frieder, Lionel Wong, Miri Zilka, **Umang Bhatt**, Thomas Lukasiewicz,
 Yuhuai Wu, Joshua Tenenbaum, William Hart, Timothy Gowers, Wenda Li, Adrian Weller, Mateja Jamnik
- [6] **Large Language Models Must Be Taught to Know What They Don't Know**
Neural Information Processing Systems (NeurIPS), 2024.
 Sanyam Kapoor, Nate Gruver, Manley Roberts, Katherine Collins, Arka Pal, **Umang Bhatt**, Adrian
 Weller, Samuel Dooley, Micah Goldblum, Andrew Gordon Wilson
- [7] **Perspectives on Incorporating Expert Feedback into Model Updates**
Patterns, 2023.
 Valerie Chen*, **Umang Bhatt***, Hoda Heidari, Adrian Weller, Ameet Talwalkar
- [8] **Algorithmic Loafing and Mitigation Strategies in Human-AI Teams**
Computers in Human Behavior: Artificial Humans, 2023.
 Isa Inuwa-Dutse, Alice Toniolo, Adrian Weller, **Umang Bhatt**
- [9] **Selective Concept Models: Permitting Stakeholder Customisation at Test-Time**
AAAI Conference on Human Computation and Crowdsourcing (HCOMP) 2023
 Matthew Barker, Katherine Collins, Krishnamurthy Dvijotham, Adrian Weller, **Umang Bhatt**
- [10] **FeedbackLogs: Recording and Incorporating Stakeholder Feedback into Machine Learning Pipelines**
ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO) 2023
 Matthew Barker, Emma Kallina, Dhananjay Ashok, Katherine Collins, Ashley Casovan, Adrian
 Weller, Ameet Talwalkar, Valerie Chen, **Umang Bhatt**
- [11] **Human-in-the-Loop Mixup**
Uncertainty in Artificial Intelligence (UAI) 2023 (Oral Presentation)
 Katherine Collins, **Umang Bhatt**, Weiyang Liu, Vihari Piratla, Ilia Sucholutsky, Bradley Love, Adrian Weller
- [12] **How Transparency Modulates Trust in Artificial Intelligence**
Patterns, 2022.
 John Zerilli, **Umang Bhatt**, Adrian Weller
- [13] **Eliciting and Learning with Soft Labels from Every Annotator**
AAAI Conference on Human Computation and Crowdsourcing (HCOMP) 2022
 Katherine Collins*, **Umang Bhatt***, Adrian Weller
- [14] **On the Utility of Prediction Sets in Human-AI Teams**
International Joint Conference on Artificial Intelligence (IJCAI) 2022 (Oral Presentation)
 Varun Babbar, **Umang Bhatt**, Adrian Weller
- [15] **On the Fairness of Causal Algorithmic Recourse**
AAAI Conference on Artificial Intelligence (AAAI) 2022 (Oral Presentation)
 Julius von Kügelgen, Amir Karimi, **Umang Bhatt**, Isabel Valera, Adrian Weller, Bernhard Schölkopf
- [16] **Uncertainty as a Form of Transparency: Measuring and Communicating Uncertainty**
AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES) 2021
Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q. Vera Liao, Prasanna Sattigeri, Riccardo Fogliato,
 Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, Lama Nachman, Rumi
 Chunara, Madhulika Srikumar, Adrian Weller, Alice Xiang
- [17] **Getting a CLUE: A Method for Explaining Uncertainty Estimates**
International Conference on Learning Representations (ICLR) 2021 (Oral Presentation)
 Javier Antorán, **Umang Bhatt**, Tameem Adel, Adrian Weller, José Miguel Hernández-Lobato

- [18] **Evaluating and Aggregating Feature-based Explanations**
International Joint Conference on Artificial Intelligence (IJCAI) 2020
Umang Bhatt, Adrian Weller, José M.F. Moura
- [19] **Explainable Machine Learning in Deployment**
ACM Conference on Fairness, Accountability, and Transparency (FAccT) 2020
Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José M.F. Moura, Peter Eckersley
- [20] **You Shouldn't Trust Me: Learning Models Which Conceal Unfairness From Multiple Explanation Methods**
European Conference on Artificial Intelligence (ECAI) 2020
Botty Dimanov, **Umang Bhatt**, Mateja Jamnik, Adrian Weller

TEACHING EXPERIENCE	University of Cambridge	
	• Lecturer, <i>Algorithms for Human-AI Collaboration</i>	Lent 2026
	• Lecturer, <i>Responsible Artificial Intelligence</i>	Michaelmas 2025
	African Institute on Mathematical Sciences – South Africa	
	• Visiting Lecturer, <i>Responsible Artificial Intelligence</i>	Mar 2026 – Apr 2026
	New York University	
	• Course Instructor, <i>Responsible Data Science</i> (DS-UA 202)	Spring 2024, 2025
	University of Cambridge	
	• Supervisor, <i>Inference</i> (3F8)	Lent 2022, 2023
	• Grader, <i>Probabilistic ML</i> (4F13)	Michaelmas 2021
SUPERVISION	Carnegie Mellon University	
	• Teaching Assistant, <i>Machine Learning for Engineers - Masters</i> (18-661)	Spring 2019
	• Teaching Assistant, <i>Machine Learning - PhD</i> (10-701)	Fall 2018
	• Teaching Assistant, <i>Practical Data Science</i> (15-688)	Spring 2018
	University of Cambridge	
	• Drew Calcagno, PhD	From Oct 2025
	• Elaf Almahmoud, PhD	From Oct 2025
	• Noah Broestl, PhD (co-advised with Adrian Weller)	From Oct 2025
	• Bernardo Villegas, PhD	From Oct 2025
	Alan Turing Institute	
	• Mackenzie Jorgensen, Postdoctoral Research Associate	2024 – 2025
	• Kendall Brogle, Research Assistant	2024 – 2025
	New York University	
	• Mihir Upadhyay, MS in Data Science	2024 – 2025
	• Ghina Al Shdaifat, MS in Data Science	2024 – 2025
	• Arina Shah, BA in Philosophy	2024 – 2025
	University of Cambridge	
	• Matthew Barker, MEng in Information Engineering (<i>Next: Trustwise</i>)	2022 – 2023
	• Vivek Palaniappan, MEng in Information Engineering (<i>Next: Citadel</i>)	2022 – 2023
	• Varun Babbar, MEng in Information Engineering (<i>Next: PhD at Duke</i>)	2021 – 2022
SELECT INVITED TALKS	• Javier Abad Martinez, Research Assistant (<i>Next: PhD at ETH Zurich</i>)	2021 – 2022
	• Dan Ley, MEng in Information Engineering (<i>Next: PhD at Harvard</i>)	2020 – 2021
	• ETH Zurich, <i>Computer Science Colloquium</i>	Apr 2025
	• Google, <i>Responsible AI Talk Series</i>	Jul 2024
	• KTH Royal Institute of Technology, <i>RPL Summer School</i>	Jun 2024
	• US Department of Defense, <i>CDAO Seminar</i>	Feb 2024
	• KAUST, <i>Rising Stars in AI Symposium</i>	Feb 2024
	• Instituto dos Advogados de São Paulo - IASP, <i>Seminar</i>	Dec 2023
	• House of Lords Select Committee on AI in Weapon Systems	Jun 2023

