# **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	2025 – Present 2025 – Present
	New York University Assistant Professor & Faculty Fellow, Center for Data Science	2023 - 2025
	The Alan Turing Institute Senior Research Associate, Safe and Ethical AI Team Research Associate, Safe and Ethical AI Team	2024 - 2025 2023 - 2024
	<b>Harvard University</b> Research Fellow, Center for Research on Computation and Society	2022 - 2023
Education	University of Cambridge Ph.D. in Engineering Passed with No Corrections Thesis: Trustworthy Machine Learning: From Algorithmic Transparency to Dec Committee: Adrian Weller MBE (Advisor), Neil Lawrence, Eric Horvitz	**
	Affiliations: Machine Learning Group, Computation and Biological Learning Lal Carnegie Mellon University, Pittsburgh, PA M.S. in Electrical and Computer Engineering B.S. in Electrical and Computer Engineering	2017 - 2019 2015 - 2019
SELECT FELLOWSHIPS AND AWARDS	Center for Democracy & Technology Fellowship J.P. Morgan AI PhD Fellowship (Declined)  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] <b>Documenting Deployment with Fabric: A Repository of Real-Worl</b> **AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)  **Mackenzie Jorgensen, Kendall Brogle, Katherine Collins, Lujain Ibrahim, Arina S  **Broestl, Gabriel Piles, Paul Dongha, Hatim Abdulhussein, Adrian Weller, Jillian Po	2025 Shah, Petra Ivanovic, Noah

## [2] Learning Personalized Decision Support Policies

AAAI Conference on Artificial Intelligence (AAAI) 2025 (Oral Presentation) Umang Bhatt\*, Valerie Chen\*, Katherine Collins, Parameswaran Kamalaruban, Emma Kallina, Adrian Weller, Ameet Talwalkar

# [3] Building Machines that Learn and Think with People

Nature Human Behavior, 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	University of Cambridge	
Experience	• Lecturer, Algorithms for Human-AI Collaboration	Lent 2026
	Lecturer, Responsible Artificial Intelligence	Michaelmas 2025
	African Institute on Mathematical Sciences – South Africa	
	Visiting Lecturer, Responsible Artificial Intelligence	Mar 2026 - Apr 2026
	New York University	_
	• Course Instructor, Responsible Data Science (DS-UA 202)	Spring 2024, 2025
	University of Cambridge	
	• Supervisor, Inference (3F8)	Lent 2022, 2023
	Grader, Probabilistic ML (4F13)	Michaelmas 2021
	Carnegie Mellon University	
	• Teaching Assistant, Machine Learning for Engineers - Masters (18-661)	Spring 2019
	• Teaching Assistant, Machine Learning - PhD (10-701)	Fall 2018
	Teaching Assistant, Practical Data Science (15-688)	Spring 2018
Supervision	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 (Next: Trustwise)	2022 - 2023
	• Vivek Palaniappan, MEng in Information Engineering (Next: Citadel)	2022 - 2023
	• Varun Babbar, MEng in Information Engineering (Next: PhD at Duke)	2021 - 2022
	• Javier Abad Martinez, Research Assistant (Next: PhD at ETH Zurich)	2021 - 2022
	• Dan Ley, MEng in Information Engineering (Next: PhD at Harvard)	2020 - 2021
Select	• ETH Zurich, Computer Science Colloquium	Apr 2025
Invited	Google, Responsible AI Talk Series	Jul 2024
Talks	KTH Royal Institute of Technology, RPL Summer School	Jun 2024
	US Department of Defense, CDAO Seminar	Feb 2024
	KAUST, Rising Stars in AI Symposium	Feb 2024
	<ul> <li>Instituto dos Advogados de São Paulo - IASP, Seminar</li> </ul>	Dec 2023
	TT CT 101 CO W ATT TT 0	T 0000

Jun 2023

• House of Lords Select Committee on AI in Weapon Systems

	• Indian Institute of Technology Bombay, C-MInDS Seminar	Jan 2023
	DeepMind, Ethics Research Team	July 2022
	• University of Manchester, Advances in Data Science and AI Conference	June 2022
	Ada Lovelace Institute, Brown Bag Lunch Seminar	June 2022
	Pennsylvania State University, Young Achievers Symposium	Apr 2022
	UK Ministry of Defense, AI Safety Workshop	Mar 2022
	Technical University of Denmark, <i>Trustworthiness in ML Seminar</i>	Apr 2021
	Imperial College, Explainable AI Seminar	Feb 2021
	ICML Workshop on Extending Explainable AI (Keynote)	July 2020
	QuantumBlack (McKinsey), AI Seminar	May 2020
	Quantumblack (Werlinsey), 711 Seminar	Way 2020
PROFESSIONAL	Finance Chair, ACM Conference on Fairness, Accountability, and Transparency (F.	AccT) 2024
SERVICE	Associate Chair, International Conference on Machine Learning (ICML)	2024
SERVICE	Co-Organizer	2024
	Deep Learning Indaba Workshop: Human-AI Interaction for the Global Majority	Aug 2025
	DALI Workshop: Uncertainty Communication	Apr 2025
	Deep Learning Indaba Practical: Responsible AI	Sept 2024
		Mar 2024
	ELSA Workshop: Generative AI and Creative Arts  Door I coming Indeha Prostical Responsible AI	
	Deep Learning Indaba Practical: Responsible AI  ICAN Workshop and House Marking Collaboration and Transition and Transiti	Sept 2023
	ICML Workshop: Human-Machine Collaboration and Teaming  THE Workshop: Human-Machine Collaboration and Teaming	July 2022
	ELLIS Workshop: Human-Centric Machine Learning  LONG W. L.	May 2021
	ICML Workshop: Human Interpretability in Machine Learning  The Company of th	July 2020
	IBM + Partnership on AI Workshop: Explainable AI	Feb 2020
	<ul> <li>Program Committee (Reviewer)</li> <li>2025: AIES, AISTATS, FAccT (Area Chair), ICML, NeurIPS, UAI</li> <li>2024: AIES, FAccT, NeurIPS</li> <li>2023: FAccT, ICLR, ICML, IJCAI, NeurIPS, UIST, WebConf</li> <li>2022: AAAI, AISTATS, FAccT, ICLR, ICML, NeurIPS, UAI</li> <li>2021: AAAI, AISTATS, FAccT, ICAIF, ICLR, ICML, KDD, NeurIPS, UAI</li> <li>2020: ICAIF, NeurIPS</li> </ul>	
	<b>Reviewer</b> : ACM Computing Surveys, ACM Transactions on Computer-Human Intera ACM Transactions on Interactive Intelligent Systems (TiiS), Artificial Intelligence J Harvard Data Science Review (HDSR), IEEE Transactions on Pattern Analysis and M ligence (TPAMI), Journal of Artificial Intelligence Research (JAIR), Journal of Mach Research (JMLR), Transactions on Machine Learning Research (TMLR), Management	Journal (AIJ), lachine Intel- ine Learning
Affiliations	Bourne-Epsom Protocol	
		023 – Present
	OECD	
		023 – Present
	Leverhulme Center for the Future of Intelligence	
	Associate Fellow	2023 - 2025
	Student Fellow	2019 - 2023
		2020
Отнек	Venture Partner, <b>HOF Capital</b> 20	)25 – Present
Experience	, 1	024 – Present
	•	023 – Present
	Subject Matter Expert, <b>Accenture</b>	2023 - 2024
	Evnert Advisor Desnonsible Al Institute	2023 2021

2021 - 2023

2020 - 2021

2017 - 2018

Expert Advisor, Responsible AI Institute

Advisor, Credo AI

 $\hbox{Co-Founder, } \textbf{Percepsense}$