

Umang Bhatt

CONTACT usb20@cam.ac.uk
INFORMATION Citizenship: USA

[Website](#)
[Google Scholar](#)

EDUCATION **University of Cambridge**, Cambridge, UK
Ph.D. in Engineering Sept 2019 – Present
Advisor: [Adrian Weller](#) MBE
Affiliations: [Machine Learning Group](#), Computation and Biological Learning Lab
Carnegie Mellon University, Pittsburgh, PA
M.S. in Electrical and Computer Engineering Aug 2017 – May 2019
B.S. in Electrical and Computer Engineering Aug 2015 – May 2019

POSITIONS Research Associate, **The Alan Turing Institute** Jan 2023 – Present
Co-lead the Turing Interest Group on Human-Machine Teams
Research Fellow, **Harvard University** June 2022 – Mar 2023
Center for Research on Computation and Society
Hosts: Milind Tambe (SEAS) and Himabindu Lakkaraju (HBS)

SELECT **J.P. Morgan AI PhD Fellowship** 2022 – 2023
FELLOWSHIPS **The Alan Turing Institute Enrichment Studentship** 2021 – 2022
AND AWARDS **Mozilla Fellowship** 2020 – 2021
Partnership on AI Research Fellowship 2019 – 2020
Leverhulme Center for the Future of Intelligence PhD Scholarship 2019 – 2023
Fully funded by DeepMind and the Leverhulme Trust
Best Presentation Award, AAAI Spring Symposium on Interpretable AI for Well-Being 2019
Lovett Family Endowed Scholarship, The Andrew Carnegie Society 2019
Undergraduate Research Presentation Award, CMU 2017
NSF I-Corps Site Award for research commercialization 2017
H. F. McCullough Memorial Scholarship, CMU 2016

PEER- [1] **Human-in-the-Loop Mixup**
REVIEWED *Uncertainty in Artificial Intelligence (UAI) 2023*
CONFERENCE Katherine Collins, **Umang Bhatt**, Weiyang Liu, Vihari Piratla, Ilia Sucholutsky, Bradley Love, Adrian Weller
PUBLICATIONS [2] **On the Informativeness of Supervision Signals**
Uncertainty in Artificial Intelligence (UAI) 2023
Ilia Sucholutsky, Ruairidh McLennan Battleday, Katherine Collins, Raja Marjeh, Joshua Peterson, Pulkit Singh, **Umang Bhatt**, Nori Jacoby, Adrian Weller, Thomas Griffiths
[3] **Human Uncertainty in Concept-Based AI Systems**
AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES) 2023
Katherine Collins, Matthew Barker, Mateo Espinosa Zarlenga, Naveen Raman, **Umang Bhatt**, Mateja Jamnik, Ilia Sucholutsky, Adrian Weller, Krishnamurthy Dvijotham
[4] **Iterative Teaching by Data Hallucination**
International Conference on Artificial Intelligence and Statistics (AISTATS) 2023
Zeju Qiu, Weiyang Liu, Tim Xiao, Zhen Liu, Yucen Luo, **Umang Bhatt**, Adrian Weller, Bernhard Schölkopf
[5] **Approximating Full Conformal Prediction at Scale via Influence Functions**
AAAI Conference on Artificial Intelligence (AAAI) 2023
Javier Abad Martinez, **Umang Bhatt**, Adrian Weller, Giovanni Cherubin
[6] **Towards Robust Metrics for Concept Representation Evaluation**
AAAI Conference on Artificial Intelligence (AAAI) 2023
Mateo Zarlenga, Pietro Barbiero, Zohreh Shams, D. Kazhdan, **Umang Bhatt**, Adrian Weller, Mateja Jamnik

- [7] **Uncertainty Quantification with Pre-trained Language Models: An Empirical Analysis**
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2022 (Findings)
Yuxin Xiao, Paul Pu Liang, **Umang Bhatt**, Willie Neiswanger, Ruslan Salakhutdinov, L.P. Morency
- [8] **Eliciting and Learning with Soft Labels from Every Annotator**
AAAI Conference on Human Computation and Crowdsourcing (HCOMP) 2022
Katherine Collins*, **Umang Bhatt***, Adrian Weller
- [9] **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
- [10] **Diverse and Amortised Counterfactual Explanations for Uncertainty Estimates**
AAAI Conference on Artificial Intelligence (AAAI) 2022
Dan Ley, **Umang Bhatt**, Adrian Weller
- [11] **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
- [12] **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
- [13] **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
- [14] **FIMAP: Feature Importance by Minimal Adversarial Perturbation**
AAAI Conference on Artificial Intelligence (AAAI) 2021
Matt Chapman-Rounds, **Umang Bhatt**, Erik Pazos, Marc-Andre Schulz, Kostas Georgatzis
- [15] **Evaluating and Aggregating Feature-based Explanations**
International Joint Conference on Artificial Intelligence (IJCAI) 2020
Umang Bhatt, Adrian Weller, José M.F. Moura
- [16] **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
- [17] **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
- [18] **On Network Science and Mutual Information for Explaining Deep Neural Networks**
IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020
Brian Davis*, **Umang Bhatt***, Kartikeya Bhardwaj*, Radu Marculescu, José M.F. Moura
- [19] **A Robot's Expressive Language Affects Human Strategy and Perceptions in a Competitive Game**
IEEE Conference on Robot and Human Interactive Communication (ROMAN) 2019
Aaron Roth, Samantha Reig, **Umang Bhatt**, Johnathan Schulgach, Tamara Amin, Afsaneh Doryab, Fei Fang, Manuela Veloso
- [20] **How Transparency Modulates Trust in Artificial Intelligence**
Patterns. Cell Press 2022
John Zerilli, **Umang Bhatt**, Adrian Weller

BOOK CHAPTERS	[21] Trust in Artificial Intelligence: Clinicians are Essential <i>Healthcare Information Technology for Cardiovascular Medicine</i> . Springer 2021 Umang Bhatt , Zohreh Shams	
SELECT WORKSHOP PUBLICATIONS	[22] GeValDi: Generative Validation of Discriminative Models <i>ICLR Workshop on Pitfalls of Limited Data and Computation for Trustworthy ML 2023</i> Vivek Palaniappan, Matt Ashman, Katherine Collins, Juyeon Heo, Adrian Weller, Umang Bhatt	
TEACHING EXPERIENCE	Thesis Co-Supervisor/Mentor , University of Cambridge <ul style="list-style-type: none"> Matthew Barker, MEng in Information Engineering May 2022 – June 2023 Vivek Palaniappan, MEng in Information Engineering (<i>Next: Citadel</i>) May 2022 – June 2023 Katherine Collins, MPhil in Machine Learning (<i>Next: PhD at Cambridge</i>) Nov 2021 – Sept 2022 Received Departmental Distinction for her thesis Varun Babbar, MEng in Information Engineering (<i>Next: PhD at Duke</i>) May 2021 – June 2022 Won the 2022 Engineering Division F Thesis Prize Javier Abad Martinez, Research Assistant (<i>Next: PhD at ETH Zurich</i>) Nov 2021 – Feb 2022 Dan Ley, MEng in Information Engineering (<i>Next: PhD at Harvard</i>) May 2020 – Aug 2021 Won the 2021 Engineering Division F Thesis Prize Teaching Assistant (Supervisor/Grader) , University of Cambridge <ul style="list-style-type: none"> <i>Inference</i> (3F8) for J.M. Hernández-Lobato and David Krueger Lent 2023 <i>Inference</i> (3F8) for Richard Turner and David Krueger Lent 2022 <i>Probabilistic ML</i> (4F13) for Zoubin Ghahramani and J.M. Hernández-Lobato Michaelmas 2020 Teaching Assistant , Carnegie Mellon University <ul style="list-style-type: none"> <i>Machine Learning for Engineers - Masters</i> (18-661) for Gauri Joshi Spring 2019 <i>Machine Learning - PhD</i> (10-701) for Ziv Bar-Joseph and Pradeep Ravikumar Fall 2018 <i>Practical Data Science</i> (15-688) for Zico Kolter Spring 2018 <i>Principles of Imperative Computation</i> (15-122) for Illiano Cervesato Fall 2017 <i>Principles of Computing</i> (15-110) for Margret Reid-Miller Spring 2017 	
SELECT INVITED TALKS	<ul style="list-style-type: none"> Indian Institute of Technology Bombay, <i>C-MInDS Seminar</i> Jan 2023 University of Chicago, <i>Chicago Human+AI Lab</i> Nov 2022 Massachusetts Institute of Technology, <i>AI Ethics and Policy Group</i> Oct 2022 Birkbeck, University of London, <i>Workshop on Human Behavioral Aspects of XAI</i> Sept 2022 Harvard Business School, <i>Digital, Data, and Design (D³) Institute</i> Sept 2022 DeepMind, <i>Ethics Research Team</i> July 2022 University of Bristol, <i>REPHRAIN Masterclass</i> June 2022 University of Cambridge, <i>Guest Lecture for MSt in AI Ethics and Society</i> June 2022 University of Manchester, <i>Advances in Data Science and AI Conference</i> June 2022 Leverhulme Centre for the Future of Intelligence, <i>Seminar Series</i> June 2022 Ada Lovelace Institute, <i>Brown Bag Lunch</i> June 2022 Von Hugel Institute, <i>Research Salon on Artificial Intelligence</i> June 2022 Pennsylvania State University, <i>Young Achievers Symposium</i> Apr 2022 UK Ministry of Defense, <i>AI Safety Workshop</i> Mar 2022 Huawei, <i>Strategy & Technology Workshop</i> Oct 2021 QuantUniversity, <i>Summer School Speaker Series</i> July 2021 Vanguard Health, <i>MedTech Insight Podcast</i> June 2021 Technical University of Denmark, <i>Trustworthiness and Interpretability in ML Seminar</i> Apr 2021 Harvard SEAS, <i>Guest Lecture for COMPSCI 282BR: Explainability in ML</i> Apr 2021 Imperial College, <i>Explainable AI Seminar</i> Feb 2021 Cambridge Department of Chemical Engineering Oct 2020 Cambridge Observatory for Human-Machine Collaboration (<i>Keynote</i>) Sept 2020 Robust and Responsible AI Developers (<i>Keynote</i>) July 2020 ICML Workshop on Extending Explainable AI (<i>Keynote</i>) July 2020 Mozilla All-Hands Meeting (<i>Keynote</i>) June 2020 QuantumBlack (McKinsey), <i>AI Seminar</i> May 2020 	

	<ul style="list-style-type: none"> • Fiddler Labs, <i>Explainable AI Seminar</i> • Element AI, <i>Explainability Research Team</i> • AILA, <i>Symposium on Creating a Fair and Ethical Future</i> • AAAI, <i>Spring Symposium on AI and Society</i> • University of Chicago, <i>Data Science for Social Good Conference</i> • Bloomberg, <i>Data for Good Exchange Conference</i> 	<p>May 2019</p> <p>Dec 2018</p> <p>Oct 2018</p> <p>Mar 2018</p> <p>Oct 2017</p> <p>Sept 2017</p>
PROFESSIONAL SERVICE	<p>Co-Organizer</p> <ul style="list-style-type: none"> • ICML Workshop on Human-Machine Collaboration and Teaming • ELLIS Workshop on Human-Centric Machine Learning • ICML Workshop on Human Interpretability in Machine Learning • IBM + Partnership on AI Workshop on Explainable AI <p>Program Committee (Reviewer)</p> <ul style="list-style-type: none"> • 2023: FAccT, ICLR, ICML, IJCAI, NeurIPS, UIST, WebConf • 2022: AAAI, AISTATS, FAccT, ICLR, ICML, NeurIPS, UAI • 2021: AAAI, AISTATS, FAccT, ICAIF, ICLR, ICML, KDD, NeurIPS, UAI • 2020: ICAIF, NeurIPS <p>Reviewer: ACM Transactions on Computer-Human Interaction (TOCHI), ACM Transactions on Interactive Intelligent Systems (TiiS), Artificial Intelligence Journal (AIJ), Journal of Artificial Intelligence Research (JAIR), IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Transactions on Machine Learning Research (TMLR)</p>	<p>July 2022</p> <p>May 2021</p> <p>July 2020</p> <p>Feb 2020</p>
OTHER EXPERIENCE	<p>Advisor, OECD, Paris, FR</p> <ul style="list-style-type: none"> • Expert Group on AI Risk & Accountability <p>Advisor, Responsible AI Institute, Austin, TX</p> <ul style="list-style-type: none"> • Building tools and certification programs for Responsible AI <p>Advisor, Credo AI, Palo Alto, CA</p> <ul style="list-style-type: none"> • Scoped an AI governance and auditing platform; backed by AI Fund <p>Research Assistant, Carnegie Mellon University, Pittsburgh, PA</p> <ul style="list-style-type: none"> • Advisors: José M.F. Moura (ECE), Pradeep Ravikumar (MLD), and Zico Kolter (CSD) <p>Student Fellow, .406 Ventures, Boston, MA</p> <ul style="list-style-type: none"> • Sourced startups and performed first-round due diligence on ventures <p>Co-Founder, Perceptsense, Pittsburgh, PA</p> <ul style="list-style-type: none"> • Built products to harvest vehicular telematics data; pipeline acquired by Honda Motors 	<p>Apr 2023 – Present</p> <p>Oct 2021 – Apr 2023</p> <p>June 2020 – June 2021</p> <p>Jan 2017 – Sept 2019</p> <p>July 2017 – June 2019</p> <p>Jan 2017 – May 2018</p>