

Umang Bhatt

CONTACT umangbhatt@nyu.edu
INFORMATION Citizenship: USA

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

CURRENT Assistant Professor and Faculty Fellow, **New York University** Oct 2023 – Present
POSITIONS Center for Data Science
Research Associate, **The Alan Turing Institute** Jan 2023 – Present
Project ELSA: European Lighthouse on Secure and Safe AI

EDUCATION **University of Cambridge**, Cambridge, UK
Ph.D. in Engineering Sept 2019 – Nov 2023
Passed with No Corrections
Advisor: [Adrian Weller](#) MBE
Examiners: Neil Lawrence, Eric Horvitz
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

SELECT **Center for Democracy & Technology Fellowship** 2024 – 2026
FELLOWSHIPS **J.P. Morgan AI PhD Fellowship** (*Declined*) [🔗](#) 2022 – 2023
AND AWARDS **The Alan Turing Institute Enrichment Studentship** [🔗](#) 2021 – 2022
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] **Selective Concept Models: Permitting Stakeholder Customisation at Test-Time**
REVIEWED *AAAI Conference on Human Computation and Crowdsourcing (HCOMP) 2023*
CONFERENCE Matthew Barker, Katherine Collins, Krishnamurthy Dvijotham, Adrian Weller, **Umang Bhatt**
PUBLICATIONS [2] **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**
[3] **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
[4] **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
[5] **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

- [6] **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
- [7] **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
- [8] **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
- [9] **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
- [10] **Eliciting and Learning with Soft Labels from Every Annotator**
AAAI Conference on Human Computation and Crowdsourcing (HCOMP) 2022
 Katherine Collins*, **Umang Bhatt***, Adrian Weller
- [11] **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
- [12] **Diverse and Amortised Counterfactual Explanations for Uncertainty Estimates**
AAAI Conference on Artificial Intelligence (AAAI) 2022
 Dan Ley, **Umang Bhatt**, Adrian Weller
- [13] **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
- [14] **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
- [15] **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
- [16] **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
- [17] **Evaluating and Aggregating Feature-based Explanations**
International Joint Conference on Artificial Intelligence (IJCAI) 2020
Umang Bhatt, Adrian Weller, José M.F. Moura
- [18] **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
- [19] **You Shouldn't Trust Me: Learning Models Which Conceal Unfairness From Multiple Explanation Methods**
European Conference on Artificial Intelligence (ECAI) 2020
 Boty Dimanov, **Umang Bhatt**, Mateja Jamnik, Adrian Weller

	[20] On Network Science and Mutual Information for Explaining Deep Neural Networks <i>IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020</i> Brian Davis*, Umang Bhatt* , Kartikeya Bhardwaj*, Radu Marculescu, José M.F. Moura	
JOURNAL PAPERS	[21] Evaluating Language Models for Mathematics through Interactions Forthcoming. <i>Proceedings of the National Academy of Sciences</i> , 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	
	[22] Algorithmic Loafing and Mitigation Strategies in Human-AI Teams <i>Computers in Human Behavior: Artificial Humans</i> , 2023. Isa Inuwa-Dutse, Alice Toniolo, Adrian Weller, Umang Bhatt	
	[23] Perspectives on Incorporating Expert Feedback into Model Updates <i>Patterns</i> , 2023. Valerie Chen*, Umang Bhatt* , Hoda Heidari, Adrian Weller, Ameet Talwalkar	
	[24] How Transparency Modulates Trust in Artificial Intelligence <i>Patterns</i> , 2022. John Zerilli, Umang Bhatt , Adrian Weller	
BOOK CHAPTERS	[25] Trust in Artificial Intelligence: Clinicians are Essential <i>Healthcare Information Technology for Cardiovascular Medicine</i> , 2021 Umang Bhatt , Zohreh Shams	
PREPRINTS	[26] When Should Algorithms Resign? Umang Bhatt* , Holli Sargeant*	
	[27] Learning Personalized Decision Support Policies Umang Bhatt* , Valerie Chen*, Katherine Collins, Parameswaran Kamalaruban, Emma Kallina, Adrian Weller, Ameet Talwalkar	
TEACHING EXPERIENCE	Advisor , <i>New York University</i> • Suha Memon, BS in Data Science • Varun Deliwala, MS in Computer Science Advisor , <i>Alan Turing Institute</i> • Meirav Segal, Research Associate Course Instructor , <i>New York University</i> • Responsible Data Science (DS-UA 202) Topics: Fairness, Accountability, Transparency, and Privacy; 158 undergraduates enrolled Thesis Co-Supervisor/Mentor , <i>University of Cambridge</i> • Matthew Barker, MEng in Information Engineering (<i>Next: Trustwise</i>) • Vivek Palaniappan, MEng in Information Engineering (<i>Next: Citadel</i>) • Katherine Collins, MPhil in Machine Learning (<i>Next: PhD at Cambridge</i>) Received Departmental Distinction for her thesis • Varun Babbar, MEng in Information Engineering (<i>Next: PhD at Duke</i>) Won the 2022 Engineering Division F Thesis Prize • Javier Abad Martinez, Research Assistant (<i>Next: PhD at ETH Zurich</i>) • Dan Ley, MEng in Information Engineering (<i>Next: PhD at Harvard</i>) Won the 2021 Engineering Division F Thesis Prize	Jan 2024 – Present Jan 2024 – Present Mar 2024 – Present Spring 2024 May 2022 – June 2023 May 2022 – June 2023 Nov 2021 – Sept 2022 May 2021 – June 2022 Nov 2021 – Feb 2022 May 2020 – Aug 2021
SELECT INVITED TALKS	• US Department of Defense, <i>CDAO Seminar</i> • KAUST, <i>Rising Stars in AI Symposium</i> • Instituto dos Advogados de São Paulo - IASP, <i>Seminar</i> • House of Lords Select Committee on AI in Weapon Systems • Indian Institute of Technology Bombay, <i>C-MInDS Seminar</i> • University of Chicago, <i>Chicago Human+AI Lab</i>	Feb 2024 Feb 2024 Dec 2023 Jun 2023 Jan 2023 Nov 2022

