

## Umang Bhatt

CONTACT [usb20@cam.ac.uk](mailto:usb20@cam.ac.uk)  
 INFORMATION Citizenship: USA

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

CURRENT Assistant Professor & 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 – Sept 2023  
 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 **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] **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 Zarlega, 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**  
*IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020*  
 Brian Davis\*, **Umang Bhatt\***, Kartikeya Bhardwaj\*, Radu Marculescu, José M.F. Moura

JOURNAL PAPERS	[21] <b>Perspectives on Incorporating Expert Feedback into Model Updates</b> <i>Patterns</i> . Cell Press 2023 Valerie Chen*, <b>Umang Bhatt*</b> , Hoda Heidari, Adrian Weller, Ameet Talwalkar	
	[22] <b>How Transparency Modulates Trust in Artificial Intelligence</b> <i>Patterns</i> . Cell Press 2022 John Zerilli, <b>Umang Bhatt</b> , Adrian Weller	
BOOK CHAPTERS	[23] <b>Trust in Artificial Intelligence: Clinicians are Essential</b> <i>Healthcare Information Technology for Cardiovascular Medicine</i> . Springer 2021 <b>Umang Bhatt</b> , Zohreh Shams	
SELECT WORKSHOP PUBLICATIONS	[24] <b>GeValDi: Generative Validation of Discriminative Models</b> <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, <b>Umang Bhatt</b>	
	[25] <b>Machine Learning Explainability for External Stakeholders</b> <i>IJCAI Workshop on Explainable AI 2020</i> <b>Umang Bhatt</b> , McKane Andrus, Adrian Weller, Alice Xiang	
PREPRINTS	[26] <b>Learning Personalized Decision Support Policies</b> <b>Umang Bhatt*</b> , Valerie Chen*, Katherine Collins, Parameswaran Kamalaruban, Emma Kallina, Adrian Weller, Ameet Talwalkar	
TEACHING EXPERIENCE	<b>Thesis Co-Supervisor/Mentor</b> , University of Cambridge <ul style="list-style-type: none"> <li>Matthew Barker, MEng in Information Engineering (<i>Next: Trustwise</i>) May 2022 – June 2023</li> <li>Vivek Palaniappan, MEng in Information Engineering (<i>Next: Citadel</i>) May 2022 – June 2023</li> <li>Katherine Collins, MPhil in Machine Learning (<i>Next: PhD at Cambridge</i>) Nov 2021 – Sept 2022 Received Departmental Distinction for her thesis</li> <li>Varun Babbar, MEng in Information Engineering (<i>Next: PhD at Duke</i>) May 2021 – June 2022 Won the 2022 Engineering Division F Thesis Prize</li> <li>Javier Abad Martinez, Research Assistant (<i>Next: PhD at ETH Zurich</i>) Nov 2021 – Feb 2022</li> <li>Dan Ley, MEng in Information Engineering (<i>Next: PhD at Harvard</i>) May 2020 – Aug 2021 Won the 2021 Engineering Division F Thesis Prize</li> </ul> <b>Teaching Assistant (Supervisor/Grader)</b> , University of Cambridge <ul style="list-style-type: none"> <li><i>Inference</i> (3F8) for J.M. Hernández-Lobato and David Krueger Lent 2023</li> <li><i>Inference</i> (3F8) for Richard Turner and David Krueger Lent 2022</li> <li><i>Probabilistic ML</i> (4F13) for Zoubin Ghahramani and J.M. Hernández-Lobato Michaelmas 2020</li> </ul> <b>Teaching Assistant</b> , Carnegie Mellon University <ul style="list-style-type: none"> <li><i>Machine Learning for Engineers - Masters</i> (18-661) for Gauri Joshi Spring 2019</li> <li><i>Machine Learning - PhD</i> (10-701) for Ziv Bar-Joseph and Pradeep Ravikumar Fall 2018</li> <li><i>Practical Data Science</i> (15-688) for Zico Kolter Spring 2018</li> <li><i>Principles of Imperative Computation</i> (15-122) for Illiano Cervesato Fall 2017</li> <li><i>Principles of Computing</i> (15-110) for Margret Reid-Miller Spring 2017</li> </ul>	
SELECT INVITED TALKS	<ul style="list-style-type: none"> <li>House of Lords Select Committee on AI in Weapon Systems Jun 2023</li> <li>Indian Institute of Technology Bombay, <i>C-MInDS Seminar</i> Jan 2023</li> <li>University of Chicago, <i>Chicago Human+AI Lab</i> Nov 2022</li> <li>Massachusetts Institute of Technology, <i>AI Ethics and Policy Group</i> Oct 2022</li> <li>Birkbeck, University of London, <i>Workshop on Human Behavioral Aspects of XAI</i> Sept 2022</li> <li>DeepMind, <i>Ethics Research Team</i> July 2022</li> <li>University of Bristol, <i>REPHRAIN Masterclass</i> June 2022</li> <li>University of Cambridge, <i>Guest Lecture for MSt in AI Ethics and Society</i> June 2022</li> <li>University of Manchester, <i>Advances in Data Science and AI Conference</i> June 2022</li> <li>Leverhulme Centre for the Future of Intelligence, <i>Seminar Series</i> June 2022</li> <li>Ada Lovelace Institute, <i>Brown Bag Lunch</i> June 2022</li> <li>Von Hugel Institute, <i>Research Salon on Artificial Intelligence</i> June 2022</li> </ul>	

	<ul style="list-style-type: none"> <li>• Pennsylvania State University, <i>Young Achievers Symposium</i></li> <li>• UK Ministry of Defense, <i>AI Safety Workshop</i></li> <li>• Huawei, <i>Strategy &amp; Technology Workshop</i></li> <li>• QuantUniversity, <i>Summer School Speaker Series</i></li> <li>• Vanguard Health, <i>MedTech Insight Podcast</i></li> <li>• Technical University of Denmark, <i>Trustworthiness and Interpretability in ML Seminar</i></li> <li>• Harvard SEAS, <i>Guest Lecture for COMPSCI 282BR: Explainability in ML</i></li> <li>• Imperial College, <i>Explainable AI Seminar</i></li> <li>• Robust and Responsible AI Developers (<i>Keynote</i>)</li> <li>• ICML Workshop on Extending Explainable AI (<i>Keynote</i>)</li> <li>• Mozilla All-Hands Meeting (<i>Keynote</i>)</li> <li>• QuantumBlack (McKinsey), <i>AI Seminar</i></li> <li>• Fiddler Labs, <i>Explainable AI Seminar</i></li> <li>• Element AI, <i>Explainability Research Team</i></li> <li>• AILA, <i>Symposium on Creating a Fair and Ethical Future</i></li> <li>• AAAI, <i>Spring Symposium on AI and Society</i></li> <li>• University of Chicago, <i>Data Science for Social Good Conference</i></li> <li>• Bloomberg, <i>Data for Good Exchange Conference</i></li> </ul>	<p>Apr 2022</p> <p>Mar 2022</p> <p>Oct 2021</p> <p>July 2021</p> <p>June 2021</p> <p>Apr 2021</p> <p>Apr 2021</p> <p>Feb 2021</p> <p>July 2020</p> <p>July 2020</p> <p>June 2020</p> <p>May 2020</p> <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><b>Co-Organizer</b></p> <ul style="list-style-type: none"> <li>• <a href="#">ICML Workshop on Human-Machine Collaboration and Teaming</a></li> <li>• <a href="#">ELLIS Workshop on Human-Centric Machine Learning</a></li> <li>• <a href="#">ICML Workshop on Human Interpretability in Machine Learning</a></li> <li>• IBM + Partnership on AI Workshop on Explainable AI</li> </ul> <p><b>Program Committee (Reviewer)</b></p> <ul style="list-style-type: none"> <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> <p><b>Reviewer:</b> ACM Transactions on Computer-Human Interaction (TOCHI), ACM Transactions on Interactive Intelligent Systems (TiiS), Artificial Intelligence Journal (AIJ), IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Journal of Artificial Intelligence Research (JAIR), Journal of Machine Learning Research (JMLR), 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>Advisory Panel, <a href="#">Bourne-Epsom Protocol</a>, London, UK</p> <ul style="list-style-type: none"> <li>• Expert guidance on the technical aspects of AI in education</li> </ul> <p>Advisor, <b>OECD</b>, Paris, FR</p> <ul style="list-style-type: none"> <li>• Expert Group on AI Risk &amp; Accountability</li> </ul> <p>Expert Advisor, <b>Accenture</b>, New York, NY</p> <ul style="list-style-type: none"> <li>• Advising on Responsible AI strategy</li> </ul> <p>Research Fellow, <b>Harvard University</b></p> <p>Center for Research on Computation and Society</p> <p>Advisor, <b>Responsible AI Institute</b>, Austin, TX</p> <ul style="list-style-type: none"> <li>• Building a scalable certification program for AI systems</li> </ul> <p>Advisor, <b>Credo AI</b>, Palo Alto, CA</p> <ul style="list-style-type: none"> <li>• Scoped an AI governance and auditing platform; backed by AI Fund</li> </ul> <p>Research Assistant, <b>Carnegie Mellon University</b>, Pittsburgh, PA</p> <ul style="list-style-type: none"> <li>• Advisors: José M.F. Moura (ECE), Pradeep Ravikumar (MLD), and Zico Kolter (CSD)</li> </ul> <p>Co-Founder, <b>Perceptense</b>, Pittsburgh, PA</p> <ul style="list-style-type: none"> <li>• Built products to harvest vehicular telematics data; pipeline acquired by <b>Honda Motors</b></li> </ul>	<p>Jul 2023 – Present</p> <p>Apr 2023 – Present</p> <p>Jan 2023 – Present</p> <p>June 2022 – Mar 2023</p> <p>Oct 2021 – Present</p> <p>June 2020 – June 2021</p> <p>Jan 2017 – Sept 2019</p> <p>Jan 2017 – May 2018</p>