

# Umang Bhatt

---

|                                       |   |                       |
|---------------------------------------|---|-----------------------|
| CONTACT INFORMATION                   | <a href="mailto:usb20@cam.ac.uk">usb20@cam.ac.uk</a><br><a href="https://umangsbhatt.github.io">umangsbhatt.github.io</a>   | Citizenship: USA      |
| EDUCATION                             | <b>University of Cambridge</b> , Cambridge, UK<br>Ph.D. in Engineering (Machine Learning)<br><i>Advisor: Adrian Weller</i><br><i>Affiliations: Machine Learning Group, Computation and Biological Learning Lab</i>  | Sept 2019 – Present   |
|                                       | <b>Carnegie Mellon University</b> , Pittsburgh, PA<br>M.S. in Electrical and Computer Engineering<br><i>Advisor: José M.F. Moura</i>  | Aug 2017 – May 2019   |
|                                       | B.S. in Electrical and Computer Engineering   | Aug 2015 – May 2019   |
| POSITIONS                             | Summer Fellow, <b>Harvard Center for Research on Computation and Society</b><br><i>Hosts: Milind Tambe (SEAS) and Himabindu Lakkaraju (HBS)</i>   | From June 2022        |
|                                       | Enrichment Student, <b>The Alan Turing Institute</b>  | Oct 2021 – Present    |
|                                       | Fellow, <b>Mozilla Foundation</b>   | Oct 2020 – Dec 2021   |
|                                       | Research Fellow, <b>Partnership on AI</b>   | June 2019 – Sept 2020 |
|                                       | Research Assistant, <b>Carnegie Mellon University</b><br><i>Mentors: Pradeep Ravikumar (MLD), Zico Kolter (CSD), Fei Fang (ISR), and Radu Marculescu (ECE)</i>  | Jan 2017 – Sept 2019  |
| SELECT FELLOWSHIPS AND AWARDS         | <b>J.P. Morgan AI PhD Fellowship</b>   | 2022 – 2023           |
|                                       | <b>The Alan Turing Institute Enrichment Studentship</b>    | 2021 – 2022           |
|                                       | <b>Mozilla Fellowship</b>   | 2020 – 2021           |
|                                       | <b>Partnership on AI Research Fellowship</b>  | 2019 – 2020           |
|                                       | <b>Leverhulme Center for the Future of Intelligence PhD Scholarship</b><br>Fully funded by DeepMind and the Leverhulme Trust  | 2019 – 2023           |
|                                       | <b>Best Presentation Award</b> , AAAI Spring Symposium on Interpretable AI for Well-Being   | 2019                  |
|                                       | <b>Lovett Family Endowed Scholarship</b> , The Andrew Carnegie Society  | 2019                  |
|                                       | <b>Undergraduate Research Presentation Award</b> , CMU  | 2017                  |
|                                       | <b>NSF I-Corps Site Award</b> for research commercialization  | 2017                  |
|                                       | <b>H. F. McCullough Memorial Scholarship</b> , CMU  | 2016                  |
| PEER-REVIEWED CONFERENCE PUBLICATIONS | [1] <b>On the Utility of Prediction Sets in Human-AI Teams</b><br><i>International Joint Conference on Artificial Intelligence (IJCAI) 2022 (Long Presentation)</i><br>Varun Babbar, <b>Umang Bhatt</b> , Adrian Weller   |                       |
|                                       | [2] <b>Diverse and Amortised Counterfactual Explanations for Uncertainty Estimates</b><br><i>AAAI International Conference on Artificial Intelligence (AAAI) 2022</i><br>Dan Ley, <b>Umang Bhatt</b> , Adrian Weller  |                       |
|                                       | [3] <b>On the Fairness of Causal Algorithmic Recourse</b><br><i>AAAI International Conference on Artificial Intelligence (AAAI) 2022 (Oral Presentation)</i><br>Julius von Kügelgen, Amir Karimi, <b>Umang Bhatt</b> , Isabel Valera, Adrian Weller, Bernhard Schölkopf   |                       |
|                                       | [4] <b>Uncertainty as a Form of Transparency: Measuring and Communicating Uncertainty</b><br><i>AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES) 2021</i><br><b>Umang Bhatt</b> , 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 |                       |
|                                       | [5] <b>Getting a CLUE: A Method for Explaining Uncertainty Estimates</b><br><i>International Conference on Learning Representations (ICLR) 2021 (Oral Presentation)</i><br>Javier Antorán, <b>Umang Bhatt</b> , Tameem Adel, Adrian Weller, José Miguel Hernández-Lobato  |                       |

- [6] **FIMAP: Feature Importance by Minimal Adversarial Perturbation**  
*AAAI International Conference on Artificial Intelligence (AAAI) 2021*  
 Matt Chapman-Rounds, **Umang Bhatt**, Erik Pazos, Marc-Andre Schulz, Kostas Georgatzis
- [7] **Evaluating and Aggregating Feature-based Explanations**  
*International Joint Conference on Artificial Intelligence (IJCAI) 2020*  
**Umang Bhatt**, Adrian Weller, José M.F. Moura
- [8] **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
- [9] **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
- [10] **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
- [11] **Building Human-Machine Trust via Interpretability**  
*AAAI International Conference on Artificial Intelligence (AAAI) 2019 (Student Abstract)*  
**Umang Bhatt**, Pradeep Ravikumar, José M.F. Moura
- [12] **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

JOURNAL  
PAPERS

- [13] **How Transparency Modulates Trust in Artificial Intelligence**  
*Patterns*. Cell Press 2022  
 John Zerilli, **Umang Bhatt**, Adrian Weller

BOOK  
CHAPTERS

- [14] **Trust in Artificial Intelligence: Clinicians are Essential**  
*Healthcare Information Technology for Cardiovascular Medicine*. Springer 2021  
**Umang Bhatt**, Zohreh Shams

TEACHING  
EXPERIENCE

- Thesis Co-Supervisor/Mentor**, University of Cambridge
- Katherine Collins, MPhil in Machine Learning and Machine Intelligence Nov 2021 – Present
  - Varun Babbar, MEng in Information Engineering May 2021 – Present
  - Javier Abad Martinez, Research Assistant (*Next: PhD at ETH Zurich*) Nov 2021 – Feb 2022
  - Dan Ley, MEng in Information Engineering (*Next: PhD at Harvard*) May 2020 – Aug 2021
  - Charlie R. Smith, Research Assistant (*Next: Future of Humanity Institute*) June 2021 – Aug 2021
- Teaching Assistant (Supervisor/Grader)**, University of Cambridge
- *Inference* (3F8) for Richard Turner and David Krueger Lent 2022
  - *Probabilistic ML* (4F13) for Zoubin Ghahramani and J.M. Hernández-Lobato Michaelmas 2020
- Teaching Assistant**, Carnegie Mellon University
- *Machine Learning for Engineers - Masters* (18-661) for Gauri Joshi Spring 2019
  - *Machine Learning - PhD* (10-701) for Ziv Bar-Joseph and Pradeep Ravikumar Fall 2018
  - *Practical Data Science* (15-688) for Zico Kolter Spring 2018
  - *Principles of Imperative Computation* (15-122) for Illiano Cervesato Fall 2017
  - *Principles of Computing* (15-110) for Margret Reid-Miller Spring 2017

|              |   |                       |
|--------------|---|-----------------------|
| SELECT       | • UK Ministry of Defense, <i>AI Safety Workshop</i>   | Mar 2022              |
| INVITED      | • Huawei, <i>Strategy &amp; Technology Workshop</i>   | Oct 2021              |
| TALKS        | • Technical University of Denmark, <i>Trustworthiness and Interpretability in ML Seminar</i>    | Apr 2021              |
|              | • Harvard SEAS ( <i>Guest Lecture</i> )   | Apr 2021              |
|              | • Imperial College, <i>Explainable AI Seminar</i>   | Feb 2021              |
|              | • 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              |
|              | • Fiddler Labs, <i>Explainable AI Seminar</i>   | May 2019              |
|              | • AILA, <i>Symposium on Creating a Fair and Ethical Future</i>                                  | Oct 2018              |
|              | • AAAI, <i>Spring Symposium on AI and Society</i>   | Mar 2018              |
|              | • University of Chicago, <i>Data Science for Social Good Conference</i>                         | Oct 2017              |
| PROFESSIONAL | <b>Co-Organizer</b>   |                       |
| SERVICE      | • <a href="#">ICML Workshop on Human-Machine Collaboration and Teaming</a>                      | July 2022             |
|              | • <a href="#">ELLIS Workshop on Human-Centric Machine Learning</a>                              | May 2021              |
|              | • <a href="#">ICML Workshop on Human Interpretability in Machine Learning</a>                   | July 2020             |
|              | • IBM + Partnership on AI Workshop on Explainable AI  | Feb 2020              |
|              | <b>Program Committee</b>  |                       |
|              | • NeurIPS – Conference on Neural Information Processing Systems                                 | 2020-2022             |
|              | • UAI – Conference on Uncertainty in Artificial Intelligence                                    | 2021, 2022            |
|              | • ICML – International Conference on Machine Learning   | 2021, 2022            |
|              | • FAccT – ACM Conference on Fairness, Accountability, and Transparency                          | 2021, 2022            |
|              | • AAAI – International Conference on Artificial Intelligence                                    | 2021, 2022            |
|              | • AISTATS – Conference on Artificial Intelligence and Statistics                                | 2021, 2022            |
|              | • ICLR – International Conference on Learning Representations                                   | 2021, 2022            |
|              | • ICAIF – ACM Conference on Artificial Intelligence and Finance                                 | 2020, 2021            |
|              | • KDD – ACM Conference on Knowledge Discovery and Data Mining                                   | 2021                  |
|              | <b>Reviewer</b>   |                       |
|              | • 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 Artificial Intelligence (TAI)  |                       |
|              | • IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)                        |                       |
|              | • Transactions on Machine Learning Research (TMLR)  |                       |
| PROFESSIONAL | Advisor, <b>Responsible AI Institute</b> , Austin, TX   | Oct 2021 – Present    |
| EXPERIENCE   | • Building tools and certification programs for Responsible AI                                  |                       |
|              | Advisor, <b>Credo AI</b> , Palo Alto, CA  | June 2020 – June 2021 |
|              | • Scoped an AI governance and auditing platform; backed by AI Fund                              |                       |
|              | Student Fellow, <b>.406 Ventures</b> , Boston, MA   | July 2018 – June 2020 |
|              | • Sourced startups and performed first-round due diligence on ventures                          |                       |
|              | Program Management Intern, <b>Microsoft</b> , Redmond, WA                                       | May 2018 – Aug 2018   |
|              | • Project: explainable conversational agents for technical hardware documentation               |                       |
|              | Co-Founder, <b>Perceptense</b> , Pittsburgh, PA   | Jan 2017 – May 2018   |
|              | • Built products to harvest vehicular telematics data; pipeline acquired by <b>Honda Motors</b> |                       |