

Umang Bhatt

CONTACT usb20@cam.ac.uk
INFORMATION umangsbhatt.github.io

Citizenship: USA

EDUCATION **University of Cambridge**, Cambridge, UK
Ph.D. in Engineering (Machine Learning) Sept 2019 – Present
Advisor: [Adrian Weller](#)
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
Advisor: [José M.F. Moura](#)
B.S. in Electrical and Computer Engineering Aug 2015 – May 2019

POSITIONS Research Fellow, **Harvard University** June 2022 – Present
Hosts: Milind Tambe (SEAS) and Himabindu Lakkaraju (HBS)
Enrichment Student, **The Alan Turing Institute** Oct 2021 – Present
Fellow, **Mozilla Foundation** Oct 2020 – Dec 2021
Research Fellow, **Partnership on AI** June 2019 – Sept 2020
Research Assistant, **Carnegie Mellon University** Jan 2017 – Sept 2019
Mentors: Pradeep Ravikumar (MLD), Zico Kolter (CSD), Fei Fang (ISR), and Radu Marculescu (ECE)

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] **On the Utility of Prediction Sets in Human-AI Teams**
REVIEWED *International Joint Conference on Artificial Intelligence (IJCAI) 2022 (Long Presentation)*
CONFERENCE Varun Babbar, **Umang Bhatt**, Adrian Weller
PUBLICATIONS [2] **Diverse and Amortised Counterfactual Explanations for Uncertainty Estimates**
AAAI Conference on Artificial Intelligence (AAAI) 2022
Dan Ley, **Umang Bhatt**, Adrian Weller
[3] **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
[4] **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
[5] **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

- [6] **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
- [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
- Matthew Barker, MEng in Information Engineering May 2022 – Present
 - Vivek Palaniappan, MEng in Information Engineering May 2022 – Present
 - Katie Collins, MPhil in Machine Learning (*Next: PhD at Cambridge*) Nov 2021 – Present
 - Anna-Maria Gueorguieva, Visiting Student (*Next: MS at Berkeley*) June 2022 – Present
 - Varun Babbar, MEng in Information Engineering (*Next: JP Morgan*) May 2021 – June 2022
 Winner of 2022 Engineering Division F Thesis Prize
 - 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
 Winner of 2021 Engineering Division F Thesis Prize
 - 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

	<ul style="list-style-type: none"> • <i>Principles of Imperative Computation</i> (15-122) for Illiano Cervesato • <i>Principles of Computing</i> (15-110) for Margret Reid-Miller 	<p>Fall 2017</p> <p>Spring 2017</p>
SELECT	<ul style="list-style-type: none"> • University of Bristol, <i>REPHRAIN Masterclass</i> 	June 2022
INVITED	<ul style="list-style-type: none"> • University of Manchester, <i>Advances in Data Science and AI Conference</i> 	June 2022
TALKS	<ul style="list-style-type: none"> • UK Ministry of Defense, <i>AI Safety Workshop</i> • Huawei, <i>Strategy & Technology Workshop</i> • Technical University of Denmark, <i>Trustworthiness and Interpretability in ML Seminar</i> • Harvard SEAS (<i>Guest Lecture</i>) • Imperial College, <i>Explainable AI Seminar</i> • Robust and Responsible AI Developers (<i>Keynote</i>) • ICML Workshop on Extending Explainable AI (<i>Keynote</i>) • Mozilla All-Hands Meeting (<i>Keynote</i>) • QuantumBlack (McKinsey), <i>AI Seminar</i> • Fiddler Labs, <i>Explainable AI Seminar</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> 	<p>Mar 2022</p> <p>Oct 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>Oct 2018</p> <p>Mar 2018</p> <p>Oct 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</p> <ul style="list-style-type: none"> • NeurIPS – Conference on Neural Information Processing Systems • UAI – Conference on Uncertainty in Artificial Intelligence • ICML – International Conference on Machine Learning • FAccT – ACM Conference on Fairness, Accountability, and Transparency • AAAI – International Conference on Artificial Intelligence • AISTATS – Conference on Artificial Intelligence and Statistics • ICLR – International Conference on Learning Representations • ICAIF – ACM Conference on Artificial Intelligence and Finance • KDD – ACM Conference on Knowledge Discovery and Data Mining <p>Reviewer</p> <ul style="list-style-type: none"> • 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) 	<p>July 2022</p> <p>May 2021</p> <p>July 2020</p> <p>Feb 2020</p> <p>2020-2022</p> <p>2021, 2022</p> <p>2021, 2022</p> <p>2021, 2022</p> <p>2021, 2022</p> <p>2021, 2022</p> <p>2021, 2022</p> <p>2020, 2021</p> <p>2021</p>
PROFESSIONAL EXPERIENCE	<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>Student Fellow, .406 Ventures, Boston, MA</p> <ul style="list-style-type: none"> • Sourced startups and performed first-round due diligence on ventures <p>Program Management Intern, Microsoft, Redmond, WA</p> <ul style="list-style-type: none"> • Project: explainable conversational agents for technical hardware documentation <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>Oct 2021 – Present</p> <p>June 2020 – June 2021</p> <p>July 2018 – June 2020</p> <p>May 2018 – Aug 2018</p> <p>Jan 2017 – May 2018</p>