

Umang Bhatt

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

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 Fellow, **Mozilla Foundation** Oct 2020 – Present
Student Fellow, **Leverhulme Center for the Future of Intelligence** Sept 2019 – Present
Research Fellow, **Partnership on AI** June 2019 – Nov 2020
Research Assistant, **Carnegie Mellon University** Jan 2017 – Sept 2019
Mentors: Pradeep Ravikumar (MLD), Zico Kolter (CSD), Fei Fang (ISR), and Radu Marculescu (ECE)

PEER-REVIEWED CONFERENCE PUBLICATIONS [1] **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, Vera Liao, Gabrielle Melançon, Prasanna Sattigeri, Ranganath Krishnan, Omesh Tickoo, Riccardo Fogliato, Jason Stanley, Madhulika Srikumar, Lama Nachman, Rumi Chunara, Adrian Weller, Alice Xiang

[2] **Getting a CLUE: A Method for Explaining Uncertainty Estimates**
International Conference on Learning Representations (ICLR) 2021 (Oral)
Javier Antorán, **Umang Bhatt**, Tameem Adel, Adrian Weller, José Miguel Hernández-Lobato

[3] **FIMAP: Feature Importance by Minimal Adversarial Perturbation**
AAAI International Conference on Artificial Intelligence (AAAI) 2021
Matt Chapman-Rounds, Erik Pazos, Marc-Andre Schulz, **Umang Bhatt**, Kostas Georgatzis

[4] **Evaluating and Aggregating Feature-based Explanations**
International Joint Conference on Artificial Intelligence (IJCAI) 2020
Umang Bhatt, Adrian Weller, José M.F. Moura

[5] **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

[6] **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

[7] **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

[8] **A Robot's Expressive Language Affects Human Strategy and Perceptions in a Competitive Game**
IEEE Conference on Robot and Human Interactive Communication (ROMAN) 2019 (Oral)

Aaron Roth, Samantha Reig, **Umang Bhatt**, Johnathan Schulgach, Tamara Amin, Afsaneh Doryab, Fei Fang, Manuela Veloso

- [9] **Building Human-Machine Trust via Interpretability**
AAAI International Conference on Artificial Intelligence (AAAI) 2019 (Student Abstract)
Umang Bhatt, Pradeep Ravikumar, José M.F. Moura
- SELECT WORKSHOP PUBLICATIONS [10] **Effects of Uncertainty on the Quality of Feature Importance Estimates**
AAAI Workshop on Explainable Agency in Artificial Intelligence 2021
Torgyn Shaikhina*, **Umang Bhatt***, Roxanne Zhang, Ana Lucic, Kostas Georgatzis, Alice Xiang, Adrian Weller
- [11] **On the Fairness of Causal Algorithmic Recourse**
NeurIPS Workshop on Algorithmic Fairness through Causality and Interpretability 2020
Julius von Kügelgen, **Umang Bhatt**, Amir-Hossein Karimi, Isabel Valera, Adrian Weller, Bernhard Schölkopf
- [12] **Counterfactual Accuracies for Alternative Models**
ICLR Workshop on Machine Learning in Real Life 2020
Umang Bhatt, Adrian Weller, Muhammad Bilal Zafar, Krishna Gummadi
- [13] **Diagnostic Model Explanations: A Medical Narrative**
AAAI Spring Symposium on Interpretable AI for Well-Being 2019 (Best Presentation Award)
Umang Bhatt, Brian Davis, José M.F. Moura
- [14] **Intelligent Pothole Detection and Road Condition Assessment**
Bloomberg Data for Good Exchange (D4GX) 2017
Umang Bhatt, Edgar Xi, Shouvik Mani, J. Zico Kolter
- BOOK CHAPTERS [15] **Trust in Artificial Intelligence: Clinicians are Essential**
Healthcare Information Technology for Cardiovascular Medicine. Ed: A.B. Bhatt. Springer 2021
Umang Bhatt, Zohreh Shams
- INVITED TALKS *Beyond Feature Importance: Explaining Uncertainty Estimates*
- International Conference on Learning Representations (ICLR) May 2021
 - Technical University of Denmark Apr 2021
 - Harvard SEAS Apr 2021
- Challenges in Deploying Explainable Machine Learning*
- Imperial College Explainable AI Seminar Feb 2021
 - Robust and Responsible AI (RsqrD) Developers Meetup July 2020
 - Keynote: ICML Workshop on Extending Explainable AI July 2020
- Explainable Machine Learning in Deployment*
- Keynote: Mozilla All-Hands June 2020
 - QuantumBlack (McKinsey) AI Seminar May 2020
 - ACM Conference on Fairness, Accountability, and Transparency (FAccT) Jan 2020
- Aggregating Feature-based Model Explanations*
- International Joint Conference on Artificial Intelligence (IJCAI) July 2020
 - Fiddler Labs May 2019
 - AAAI Spring Symposium on Interpretable AI for Well-being Mar 2019
- Maintaining the Humanity of Our Models*
- AILA Symposium on Creating a Fair and Ethical Future Oct 2018
 - AAAI Spring Symposium on AI and Society Mar 2018
- Intelligent Pothole Detection*
- University of Chicago's Data Science for Social Good Conference Oct 2017
 - Bloomberg Data for Good Exchange Sept 2017

TEACHING EXPERIENCE	Thesis Co-Supervisor , University of Cambridge	
	• Dan Ley, MEng in Information Engineering	2021
	Teaching Assistant (Supervisor/Grader) , University of Cambridge	
	• <i>Probabilistic ML</i> (4F13) for Zoubin Ghahramani and J.M. Hernández-Lobato	Michaelmas 2020
	Teaching Assistant , Carnegie Mellon University	
	• <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
PROFESSIONAL SERVICE	Co-Organizer	
	• ELLIS Workshop on Human-Centric Machine Learning	May 2021
	• ICML Workshop on Human Interpretability in Machine Learning	July 2020
	• IBM + Partnership on AI Workshop on Explainable AI	Feb 2020
	Program Committee (Reviewer)	
	• 2021: AAAI, ACM FAccT, AISTATS, ICLR, ICML, KDD, NeurIPS, UAI	
	• 2020: ACM ICAIF, NeurIPS	
	• 2019: ICLR DebugML	
FELLOWSHIPS, AWARDS, AND HONORS	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
PROFESSIONAL EXPERIENCE	Advisor, Credo AI , Palo Alto, CA	June 2020 – Present
	• Creating an AI auditing platform	
	Student Fellow, .406 Ventures , Boston, MA	July 2018 – June 2020
	• Sourced startups and performed first-round due diligence on ventures	
	Program Management Intern, Microsoft , Redmond, WA	May 2018 – Aug 2018
	• Project: explainable conversational agents for technical hardware documentation	
	Co-Founder, Perceptsense , Pittsburgh, PA	Jan 2017 – May 2018
	• Built products to harvest vehicular telematics data; pipeline acquired by Honda Motors	
	Product Management Intern, Groupon , Chicago, IL	May 2017 – Aug 2017
	• Project: personalized and targeted pricing algorithms to drive purchase frequency	
	Software Engineering Intern, InquisitHealth , River Edge, NJ	June 2016 – Sept 2016
	• Project: speaker diarization and adverse event detection in doctor-patient conversations	