

# Umang Bhatt

---

CONTACT [usb20@cam.ac.uk](mailto:usb20@cam.ac.uk)  
INFORMATION [umangsbhatt.github.io](https://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  
Enrichment Student, **The Alan Turing Institute** From Oct 2021  
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, 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
  - [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, **Umang Bhatt**, Erik Pazos, Marc-Andre Schulz, 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*

	Aaron Roth, Samantha Reig, <b>Umang Bhatt</b> , Johnathan Schulgach, Tamara Amin, Afsaneh Doryab, Fei Fang, Manuela Veloso	
	[9] <b>Building Human-Machine Trust via Interpretability</b> <i>AAAI International Conference on Artificial Intelligence (AAAI) 2019 (Student Abstract)</i> <b>Umang Bhatt</b> , Pradeep Ravikumar, José M.F. Moura	
SELECT WORKSHOP PUBLICATIONS	[10] <b><math>\delta</math>-CLUE: Diverse Sets of Explanations for Uncertainty Estimates</b> <i>ICLR Workshop on Security and Safety in Machine Learning Systems 2021</i> Dan Ley, <b>Umang Bhatt</b> , Adrian Weller	
	[11] <b>On the Fairness of Causal Algorithmic Recourse</b> <i>NeurIPS Workshop on Algorithmic Fairness through Causality and Interpretability 2020</i> Julius von Kügelgen, <b>Umang Bhatt</b> , Amir-Hossein Karimi, Isabel Valera, Adrian Weller, Bernhard Schölkopf	
	[12] <b>Counterfactual Accuracies for Alternative Models</b> <i>ICLR Workshop on Machine Learning in Real Life 2020</i> <b>Umang Bhatt</b> , Adrian Weller, Muhammad Bilal Zafar, Krishna Gummadi	
	[13] <b>Diagnostic Model Explanations: A Medical Narrative</b> <i>AAAI Spring Symposium on Interpretable AI for Well-Being 2019 (Best Presentation Award)</i> <b>Umang Bhatt</b> , Brian Davis, José M.F. Moura	
	[14] <b>Intelligent Pothole Detection and Road Condition Assessment</b> <i>Bloomberg Data for Good Exchange (D4GX) 2017</i> <b>Umang Bhatt</b> , Edgar Xi, Shouvik Mani, J. Zico Kolter	
BOOK CHAPTERS	[15] <b>Trust in Artificial Intelligence: Clinicians are Essential</b> <i>Healthcare Information Technology for Cardiovascular Medicine</i> . Ed: A.B. Bhatt. Springer 2021 <b>Umang Bhatt</b> , Zohreh Shams	
SELECT INVITED TALKS	<i>Beyond Feature Importance: Explaining Uncertainty Estimates</i> <ul style="list-style-type: none"> <li>• International Conference on Learning Representations (ICLR)</li> <li>• Technical University of Denmark</li> <li>• Harvard SEAS</li> </ul> <i>Challenges in Deploying Explainable Machine Learning</i> <ul style="list-style-type: none"> <li>• Imperial College Explainable AI Seminar</li> <li>• Robust and Responsible AI (RsqrD) Developers Meetup</li> <li>• Keynote: ICML Workshop on Extending Explainable AI</li> </ul> <i>Explainable Machine Learning in Deployment</i> <ul style="list-style-type: none"> <li>• Keynote: Mozilla All-Hands</li> <li>• QuantumBlack (McKinsey) AI Seminar</li> <li>• ACM Conference on Fairness, Accountability, and Transparency (FAccT)</li> </ul> <i>Aggregating Feature-based Model Explanations</i> <ul style="list-style-type: none"> <li>• International Joint Conference on Artificial Intelligence (IJCAI)</li> <li>• Fiddler Labs</li> <li>• AAAI Spring Symposium on Interpretable AI for Well-being</li> </ul> <i>Maintaining the Humanity of Our Models</i> <ul style="list-style-type: none"> <li>• AILA Symposium on Creating a Fair and Ethical Future</li> <li>• AAAI Spring Symposium on AI and Society</li> </ul> <i>Intelligent Pothole Detection</i> <ul style="list-style-type: none"> <li>• University of Chicago's Data Science for Social Good Conference</li> <li>• Bloomberg Data for Good Exchange</li> </ul>	May 2021 Apr 2021 Apr 2021 Feb 2021 July 2020 July 2020 June 2020 May 2020 Jan 2020 July 2020 May 2019 Mar 2019 Oct 2018 Mar 2018 Oct 2017 Sept 2017

TEACHING EXPERIENCE	<b>Thesis Co-Supervisor</b> , University of Cambridge	
	<ul style="list-style-type: none"> <li>• Dan Ley, MEng in Information Engineering</li> <li>• Timothy Ye, BS in Computer Science (Part II)</li> </ul>	2020 – Present 2020 – Present
	<b>Teaching Assistant (Supervisor/Grader)</b> , University of Cambridge	
	<ul style="list-style-type: none"> <li>• <i>Probabilistic ML</i> (4F13) for Zoubin Ghahramani and J.M. Hernández-Lobato</li> </ul>	Michaelmas 2020
	<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</li> <li>• <i>Machine Learning - PhD</i> (10-701) for Ziv Bar-Joseph and Pradeep Ravikumar</li> <li>• <i>Practical Data Science</i> (15-688) for Zico Kolter</li> <li>• <i>Principles of Imperative Computation</i> (15-122) for Illiano Cervesato</li> <li>• <i>Principles of Computing</i> (15-110) for Margret Reid-Miller</li> </ul>	Spring 2019 Fall 2018 Spring 2018 Fall 2017 Spring 2017
PROFESSIONAL SERVICE	<b>Co-Organizer</b>	
	<ul style="list-style-type: none"> <li>• ELLIS Workshop on Human-Centric Machine Learning</li> <li>• ICML Workshop on Human Interpretability in Machine Learning</li> <li>• IBM + Partnership on AI Workshop on Explainable AI</li> </ul>	May 2021 July 2020 Feb 2020
	<b>Program Committee</b>	
	<ul style="list-style-type: none"> <li>• 2021: AAAI, ACM FAccT, AISTATS, ICLR, ICML, KDD, NeurIPS, UAI</li> <li>• 2020: ACM ICAIF, NeurIPS</li> <li>• 2019: ICLR DebugML</li> </ul>	
	<b>Reviewer</b>	
	<ul style="list-style-type: none"> <li>• ACM Transactions on Computer-Human Interaction (TOCHI)</li> <li>• IEEE Transactions on Artificial Intelligence (TAI)</li> <li>• Artificial Intelligence Journal (AIJ)</li> </ul>	
FELLOWSHIPS AND AWARDS	<b>The Alan Turing Institute Enrichment Studentship</b>	2020 – 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>	2019 – 2023
	Fully funded by DeepMind and the Leverhulme Trust	
	<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
PROFESSIONAL EXPERIENCE	Advisor, <b>Credo AI</b> , Palo Alto, CA	June 2020 – Present
	<ul style="list-style-type: none"> <li>• Creating an AI auditing platform</li> </ul>	
	Student Fellow, <b>.406 Ventures</b> , Boston, MA	July 2018 – June 2020
	<ul style="list-style-type: none"> <li>• Sourced startups and performed first-round due diligence on ventures</li> </ul>	
	Program Management Intern, <b>Microsoft</b> , Redmond, WA	May 2018 – Aug 2018
	<ul style="list-style-type: none"> <li>• Project: explainable conversational agents for technical hardware documentation</li> </ul>	
	Co-Founder, <b>Perceptsense</b> , Pittsburgh, PA	Jan 2017 – May 2018
	<ul style="list-style-type: none"> <li>• Built products to harvest vehicular telematics data; pipeline acquired by <b>Honda Motors</b></li> </ul>	
	Product Management Intern, <b>Groupon</b> , Chicago, IL	May 2017 – Aug 2017
	<ul style="list-style-type: none"> <li>• Project: personalized and targeted pricing algorithms to drive purchase frequency</li> </ul>	
	Software Engineering Intern, <b>InquisitHealth</b> , River Edge, NJ	June 2016 – Sept 2016
	<ul style="list-style-type: none"> <li>• Project: speaker diarization and adverse event detection in doctor-patient conversations</li> </ul>	