

Umang Bhatt

CONTACT
INFORMATION usb20@cam.ac.uk
 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

 PhD Candidate, **University of Cambridge** Sept 2019 – Present

 Research Fellow, **Partnership on AI** May 2019 – Present

 Research Assistant, **Carnegie Mellon University** Jan 2017 – Aug 2019
 • *Mentors and Collaborators:* Pradeep Ravikumar (MLD), Zico Kolter (CSD), Fei Fang (ISR), and Radu Marculescu (ECE)

FELLOWSHIPS,
AWARDS, AND
HONORS **Mozilla Fellowship** 2020 – 2021

Leverhulme Center for the Future of Intelligence PhD Scholarship 2019 – 2022
 Fully funded by DeepMind and the Leverhulme Trust

Partnership on AI Research Fellowship 2019 – 2021

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-REVIEWED
CONFERENCE
PUBLICATIONS [1] **Evaluating and Aggregating Feature-based Explanations**
 International Joint Conference on Artificial Intelligence (IJCAI) 2020
 Umang Bhatt, Adrian Weller, José M.F. Moura

 [2] **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

 [3] **You Shouldn't Trust Me: Learning Models Which Conceal Unfairness From Multiple Explanation Methods**
 European Conference on Artificial Intelligence (ECAI) 2020
 Botby Dimanov, **Umang Bhatt**, Mateja Jamnik, Adrian Weller

 [4] **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

 [5] **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

[6] **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

[7] **On Trustworthy Artificial Intelligence for Healthcare**

Healthcare Information Technology for Cardiovascular Medicine. Ed: A. Maniar. Springer 2020

Umang Bhatt, Zohreh Shams

SELECT WORKSHOP
PUBLICATIONS

[8] **Machine Learning Explainability for External Stakeholders**

IJCAI Workshop on Explainable AI 2020

Umang Bhatt, McKane Andrus, Alice Xiang, Adrian Weller

[9] **Getting a CLUE: A Method for Explaining Uncertainty Estimates**

ICLR Workshop on Machine Learning in Real Life 2020 (Oral)

Javier Antorán, **Umang Bhatt**, Tameem Adel, Adrian Weller, José Miguel Hernández-Lobato

[10] **Counterfactual Accuracies for Alternative Models**

ICLR Workshop on Machine Learning in Real Life 2020

Umang Bhatt, Adrian Weller, Muhammad Bilal Zafar, Krishna Gummadi

[11] **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

[12] **Building Human-Machine Trust via Interpretability** (Extended Abstract)

AAAI Conference on Artificial Intelligence (AAAI) 2019

Umang Bhatt, Pradeep Ravikumar, José M.F. Moura

[13] **Maintaining the Humanity of Our Models**

AAAI Spring Symposium on AI, Society, and Ethics 2018

Umang Bhatt

INVITED TALKS

Challenges in Deploying Explainable Machine Learning

- Robust and Responsible AI (RsqrD) Developers Meetup
- Keynote: ICML Workshop on Extending Explainable AI
- QuantumBlack (McKinsey) AI seminar

July 2020

July 2020

May 2020

Explainable Machine Learning in Deployment

- Cambridge Observatory for Human-Machine Collaboration
- Keynote: Mozilla All-Hands
- ACM Conference on Fairness, Accountability, and Transparency (FAccT)

Sept 2020

June 2020

Jan 2020

Aggregating Feature-based Model Explanations

- Fiddler Labs
- Cambridge Machine Learning Group
- AAAI Spring Symposium on Interpretable AI for Well-being

May 2019

Apr 2019

Mar 2019

Neural Information Flow

- Element AI

Dec 2018

Maintaining the Humanity of Our Models

- AILA Symposium on Creating a Fair and Ethical Future
- AAAI Spring Symposium on AI and Society

Oct 2018

Mar 2018

Intelligent Pothole Detection

- University of Chicago's Data Science for Social Good Conference
- Bloomberg Data for Good Exchange

Oct 2017

Sept 2017

| | | |
|----------------------------|---|-----------------------|
| TEACHING EXPERIENCE | Thesis Co-Supervisor , University of Cambridge | |
| | • Daniel Ley, MEng Thesis | Aug 2020 – Present |
| | 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 | |
| | • ICML Workshop on Human Interpretability in Machine Learning | July 2020 |
| | • IBM + Partnership on AI Workshop on Explainable AI | Feb 2020 |
| | Program Committee | |
| | • ACM Conference on Fairness, Accountability, and Transparency (FAccT) | 2021 |
| | • Conference on Artificial Intelligence and Statistics (AISTATS) | 2021 |
| | • Conference on Artificial Intelligence (AAAI) | 2021 |
| | • International Conference on Learning Representations (ICLR) | 2021 |
| | • Conference on Neural Information Processing Systems (NeurIPS) | 2020 |
| | • ACM Conference on AI and Finance (ICAIF) | 2020 |
| | • ICLR Workshop on Debugging Machine Learning Models | 2019 |
| | Reviewer | |
| | • IEEE Transactions on Artificial Intelligence | |
| PROFESSIONAL EXPERIENCE | Advisor, Credo AI , Palo Alto, CA | June 2020 – Aug 2020 |
| | • Creating an AI auditing platform - in stealth | |
| | 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 | |