Umang Bhatt

Contact Information

usb20@cam.ac.uk
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

Enrichment Student, **The Alan Turing Institute** — Oct 2021 – Present
Student Fellow, **Leverhulme Center for the Future of Intelligence** — Sept 2019 – 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)

Peer-Reviewed Conference Publications

[1] **Diverse and Amortised Counterfactual Explanations for Uncertainty Estimates**
AAAI International Conference on Artificial Intelligence (AAAI) 2022
Dan Ley, ***Umang Bhatt***, Adrian Weller

[2] **On the Fairness of Causal Algorithmic Recourse**
AAAI International Conference on Artificial Intelligence (AAAI) 2022
Julius von Kügelgen, Amir Karimi, ***Umang Bhatt***, Isabel Valera, Adrian Weller, Bernhard Schölkopf

[3] **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

[4] **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

[5] **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

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

[7] **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

[8] **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

[9] **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

[10] **Building Human-Machine Trust via Interpretability**
AAAI International Conference on Artificial Intelligence (AAAI) 2019 (Student Abstract)
Umang Bhatt, Pradeep Ravikumar, José M.F. Moura

[11] **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 Articles

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

Select Workshop Publications

[13] **Do Concept Bottleneck Models Learn As Intended?**
ICLR Workshop on Responsible AI 2021
Andrei Margeloiu*, Matt Ashman*, ***Umang Bhatt****, Yanzhi Chen, Mateja Jamnik, Adrian Weller

[14] **Counterfactual Accuracies for Alternative Models**
ICLR Workshop on Machine Learning in Real Life 2020
Umang Bhatt, Adrian Weller, Muhammad Bilal Zafar, Krishna Gummadi

Book Chapters

[15] **Challenges in Deploying Explainable Machine Learning**
xxAI – Beyond explainable Artificial Intelligence. Springer 2022
Umang Bhatt, Alice Xiang, Shubham Sharma, Joydeep Ghosh, Ruchir Puri, José M.F. Moura, Peter Eckersley, Adrian Weller

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

Select Fellowships and Awards

J.P. Morgan AI PhD Fellowship	2022 – 2023
The Alan Turing Institute Enrichment Studentship	2021 – 2022
Mozilla Fellowship	2020 – 2021
Partnership on AI Research Fellowship	2019 – 2020
Leverhulme Center for the Future of Intelligence PhD Scholarship Fully funded by DeepMind and the Leverhulme Trust	2019 – 2023
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

Grants

Co-Investigator, "Social Explainability (SOXAI) for Trustworthy AI" 2021–2022
£107,000 from EPSRC via Research Centre on Privacy, Harm Reduction, and Adversarial Influence
PI: Frens Kroeger (Coventry); Other CIs: James Hancock (Stanford) and Beate Grawemeyer (Coventry)

Select Invited Talks

Beyond Feature Importance: Explaining Uncertainty Estimates

- Huawei Strategy & Technology Workshop — Oct 2021
- 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
- Keynote: Mozilla All-Hands Meeting 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
- Element AI Dec 2018

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 Assistantship Nov 2021 – Feb 2022
- Dan Ley, MEng in Information Engineering *(Next: JP Morgan)* May 2020 – Aug 2021
- Charlie Rogers Smith, Research Assistantship *(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 Gharamani 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

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

- 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
- NeurIPS – Conference on Neural Information Processing Systems 2020, 2021
- ICAIF – ACM Conference on Artificial Intelligence and Finance 2020, 2021
- KDD – ACM Conference on Knowledge Discovery and Data Mining 2021

Reviewer

- Journal of Artificial Intelligence Research (JAIR)
- ACM Transactions on Computer-Human Interaction (TOCHI)
- Artificial Intelligence Journal (AIJ)
- ACM Transactions on Interactive Intelligent Systems (TiiS)

Professional Experience

Advisor, **Responsible AI Institute**, Austin, TX Oct 2021 – Present

- Building tools and certification programs for Responsible AI

Advisor, **Credo AI**, Palo Alto, CA June 2020 – June 2021

- Scoped an AI governance and auditing platform; backed by AI Fund

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, **Percepsense**, Pittsburgh, PA Jan 2017 – May 2018

- Built products to harvest vehicular telematics data; pipeline acquired by ***Honda Motors***