Umang Bhatt

✉ usb20@cam.ac.uk
umangsbhatt.github.io

Education

2019 – 2023 **University of Cambridge**, Cambridge, UK.
Ph.D., Engineering (Machine Learning)
Supervisor: Adrian Weller

2017 – 2019 **Carnegie Mellon University**, Pittsburgh, PA.
M.S., Electrical and Computer Engineering
Advisor: José Moura

2015 – 2019 **Carnegie Mellon University**, Pittsburgh, PA.
B.S., Electrical and Computer Engineering

Courses Advanced Probabilistic Machine Learning (CAM), Machine Learning (CMU), Advances in AI for Social Good (CMU), Network Science (CMU), Pattern Recognition Theory (CMU)

Publications

Conference Papers

[C7] **Umang Bhatt**, Adrian Weller, & José Moura. Evaluating and Aggregating Feature-based Model Explanations. International Joint Conference on Artificial Intelligence (IJCAI) 2020. Yokohama, Japan.

[C6] **Umang Bhatt**, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José Moura, & Peter Eckersley. Explainable Machine Learning in Deployment. ACM Conference on Fairness, Accountability, and Transparency (FAT*) 2020. Barcelona, Spain.

[C5] Botty Dimanov, **Umang Bhatt**, Mateja Jamnik, & Adrian Weller. You Shouldn't Trust Me: Learning Models Which Conceal Unfairness From Multiple Explanation Methods. European Conference on Artificial Intelligence (ECAI) 2020. Santiago de Compostela, Spain.

[C4] Brian Davis*, **Umang Bhatt***, Kartikeya Bhardwaj*, Radu Marculescu, & José Moura. On Network Science and Mutual Information for Explaining Deep Neural Networks. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2020. Barcelona, Spain.

[C3] Aaron Roth, Samantha Reig, **Umang Bhatt**, Johnathan Schulgach, Tamara Amin, Afsaneh Doryab, Fei Fang, & Manuela Veloso. A Robot's Expressive Language Affects Human Strategy and Perceptions in a Competitive Game. IEEE International Conference on Robot and Human Interactive Communication (IEEE-ROMAN) 2019. New Delhi, India.

[C2] **Umang Bhatt**, Pradeep Ravikumar, & José Moura. Building Human-Machine Trust via Interpretability. Extended Abstract. AAAI 2019. Honolulu, HI.

[C1] **Umang Bhatt**, Edgar Xi, Shouvik Mani, & Zico Kolter. Intelligent Pothole Detection and Road Condition Assessment. Bloomberg Data for Good Exchange 2017. New York, NY.

Select Workshop Papers

[W4] Javier Antorán, **Umang Bhatt**, Tameem Adel, Adrian Weller, & José Miguel Hernández-Lobato. Getting a CLUE: A Method for Explaining Uncertainty Estimates. ICLR Workshop on Machine Learning in Real Life (ML-IRL), 2020. Addis Ababa, Ethiopia. **(Selected Oral Presentation)**

[W3] **Umang Bhatt**, Adrian Weller, Muhammad Bilal Zafar, & Krishna Gummadi. Counterfactual Accuracies for Alternative Models. ICLR Workshop on Machine Learning in Real Life (ML-IRL), 2020. Addis Ababa, Ethiopia.

[W2] **Umang Bhatt**, Brian Davis, & José Moura. Diagnostic Model Explanations: A Medical Narrative. AAAI Spring Symposium on Interpretable AI for Well-being, 2019. Stanford, CA. **(Best Paper Award)**

[W1] **Umang Bhatt**. Maintaining the Humanity of Our Models. AAAI Spring Symposium on AI, Society, and Ethics, 2018. Stanford, CA.

Selected Talks

2020 Invited Talk at QuantumBlack (McKinsey) AI Seminar. May 2020. London, UK.

2019 Talk at the All Partners Meeting for the Partnership on AI. September 2019. London, UK.

2019 Invited Talk at Fiddler Labs. May 2019. Palo Alto, CA.

2019 Invited Talk at Cambridge MLG. April 2019. Cambridge, UK.

2018 Invited Talk and Moderator at AI LA's AI Ethics Symposium. October 2018. Pasadena, CA.

2017 Invited Case Study at University of Chicago's DSSG. September 2017. Chicago, IL.

Academic Positions

2019 – now **Partnership on AI**, *Research Fellow*, San Francisco, CA.
Exploring how to deploy explainable machine learning in industry

2019 – now **Leverhulme Center for the Future of Intelligence**, *Student Fellow*, Cambridge, UK.
Justifying the need for trust and transparency in AI

2019 – now **University of Cambridge**, *Ph.D. Candidate*, Cambridge, UK.
Computational and Biological Learning Lab

2017 – 2019 **Carnegie Mellon University**, *Research Assistant*, Pittsburgh, PA.
Collaborated with José Moura (ECE), Pradeep Ravikumar (MLD), Radu Marculescu (ECE), Fei Fang (ISR), Zico Kolter (CSD), and Manuela Veloso (MLD)

Teaching Experience

2017 – 2019 **Carnegie Mellon University**, *Teaching Assistant*, Pittsburgh, PA.

S19 18-661 (Machine Learning for Engineers - Masters) taught by Gauri Joshi

F18 10-701 (Machine Learning - PhD) taught by Ziv Bar-Joseph and Pradeep Ravikumar

S18 15-388/15-688 (Practical Data Science) taught by Zico Kolter

F17 15-122 (Principles of Imperative Computation) taught by Illiano Cervesato

S17 15-110 (Principles of Computing) taught by Margret Reid-Miller

Service

2020 **ICML Workshop on Human Interpretability**, *Organizer*.

2019 **NeurIPS, ICML**, *Reviewer*.

2019 **ICLR Workshop on Debugging Machine Learning Models**, *Program Committee*.

2019 **AAAI/ACM Conference on AI, Ethics and Society**, *Student Programs Committee*.

2017 – 2019 **NavTalent (now HumanCapital)**, *Campus Director*.
Connected top-tier engineers with high-growth, impact-driven startups

2011 – 2019 **BAPS Children's Activities**, *National Development Committee*.
Plan, oversee, and speak at nation-wide conventions and weekly assemblies for over 5,000 Hindu children ages 6-14 with the team of fifteen other volunteers

Other Experience

2020 – now **Credo AI**, *Consultant*, Palo Alto, CA.
Creating an AI auditing platform - in stealth

2018 – 2020 **.406 Ventures**, *Student Fellow*, Boston, MA.
Sourcing startups and performing first-round due diligence on ventures

2017 – 2018 **Percepsense**, *Co-Founder*, Pittsburgh, PA.
Built products to harvest vehicular telematics data - pipeline now used by Honda Motors

2018 **Microsoft**, *Program Management Intern*, Redmond, WA.
Project: explainable conversational agents for technical hardware documentation

2017 **Groupon**, *Product Management Intern*, Chicago, IL.
Project: personalized and targeted pricing algorithms to drive purchase frequency

Skills

Python, Java, C, C#, R, Javascript, Tensorflow, PyTorch, Matlab, HTML/CSS