Finale Doshi-Velez

John A. Paulson School of Engineering and Applied Sciences, Harvard University,
33 Oxford Street Cambridge MA 02138 - **finale@seas.harvard.edu** - **http://finale.seas.harvard.edu/**

Education

- **Doctor of Philosophy**, Computer Science.
  Massachusetts Institute of Technology, September 2005 – June 2012
  Thesis: Bayesian Nonparametric Methods for Reinforcement Learning in Partially Observable Domains.
- **Master of Science**, Engineering.
  University of Cambridge, October 2007 – August 2009.
  Thesis: The Indian Buffet Process: Scalable Inference and Extensions.
- **Master of Science**, Computer Science.
  Massachusetts Institute of Technology, September 2005 – June 2007
  Thesis: Efficient Model Learning for Dialog Management
- **Bachelor of Science**, Aerospace Engineering with a minor in Creative Writing
  Bachelor of Science, Physics
  Massachusetts Institute of Technology, September 2001 – June 2005

Professional Experience

- **Gordon McKay Professor of Computer Science** July 2021 – Present
  John L. Loeb Associate Professor of Computer Science July 2019 – June 2021
  Assistant Professor of Computer Science July 2014 – June 2019
  Harvard John A. Paulson School of Engineering and Applied Sciences, Computer Science Area. My group develops machine learning methods to advance healthcare.
  Secondary affiliations: Brigham and Women's Hospital; Boston Children's Hospital; Veteran's Administration; Berkman Klein Center
- **Postdoctoral Research Associate** August 2012 – June 2014
  Center for Biomedical Informatics, Harvard Medical School. Developed machine learning methodologies to combine clinical data and expert-curated information to derive data-driven phenotypes of disease.
  Secondary affiliations: Fellow, School of Engineering and Applied Sciences at Harvard University

Selected Honors

- **Anita Borg Early Career Award** (2021)
- **Sloan Fellowship** (2018)
- **NSF CAREER Award** (2018)
- **Air Force Young Investigator Program Award** (2016)
- **IEEE's AI "10 to Watch"** (2013)
- **Marshall Scholar** (2007)

Selected Highlights

- **Machine Learning for Healthcare Conference**: Co-founder, Organizer, and Board President, 2016-present. Launched peer-reviewed, archival conference that has become the center of the ML for health community.
- **Workshop for Women in Machine Learning**: Many roles: Senior Advisory Board, Executive Board Co-Founder and President, Organizer, 2008-present. WiML provides technical feedback and networking for women in machine learning. I managed NSF grants and organized multi-year financial and admin infrastructure.
- **Mentoring Awards:** Harvard Graduate Student Council Everett Mendelsohn Excellence in Mentoring Award 2019; Harvard SEAS Capers and Marion McDonald Mentoring Award 2019
- **Graduated:** 3 graduate students and 4 postdocs; publications with 15 undergraduate students.

Publications[1]

Journal

- M. Jacobs, *M. Pradier*, M. McCoy, R. Perlis**, F. Doshi-Velez,** and G. Krzysztof, "How machine learning recommendations influence clinician treatment selections: example of antidepressant selection," *Translational Psychiatry*, vol.1, pp 1-9, 2021.
- *J. Futoma*, M. Simons, T. Panch, **F. Doshi-Velez**, L. Celi, "The Myth of Generalizability in Clinical Research and Machine Learning in Healthcare," *Lancet Digital Health*, vol. 2(9). September 2020.
- *M. Pradier, M. Hughes*, T. McCoy, S. Barroihet, **F. Doshi-Velez,** R. Perlis, "Predicting Change in Diagnosis from Major Depression to Biopolar Disorder after Antidepressant Initiation," *Neuropsychopharamcology*, to appear.
- *M. Hughes, M. Pradier, A. Ross*, T. McCoy, R. Perlis, and **F. Doshi-Velez**, "Assessment of a Prediction Model for Antidepressant Treatment Stability Using Supervised Topic Models," *JAMA Network Open*, vol. 3(5), May 2020.
- S. Ghosh, and **F. Doshi-Velez**, "Discussions on Horseshoe Regularisation for Machine Learning in Complex and Deep Models," *International Statistical Review*, vol. 1, pp.1-3, April 2020.
- **F. Doshi-Velez**, and R. Perlis, "Evaluating Machine Learning Articles," *JAMA*, vol. 322, pp 1777-1779, April 2020.
- *M. Pradier*, T. McCoy, *M. Hughes*, R. Perlis, and **F. Doshi-Velez**, "Predicting treatment dropout after antidepressant initiation," *Translational Psychiatry*, vol.10, pp 1-8, February 2020.
- McMahon, W. Cooper, J. Brown, B. Carleton, **F. Doshi-Velez**, I. Kohane, J. Goldman, M. Hoffman, R. Kamaleswaran, M. Sakiyama, S. Sekine, M. Sturkenboom, M. Turner, and R. Califf, "Big Data in the Assessment of Pediatric Medication Safety," *Pediatrics*, vol. 145 (2), pp 1-11, January 2020.
- S. Ghosh, *J. Yao*, and **F. Doshi-Velez**, "Model Selection in Bayesian Neural Networks via Horseshoe Priors," *Journal of Machine Learning Research*, vol. 20 (182), pp.1-46, October 2019.
- O. Amir, **F. Doshi-Velez**, and D. Sarne, "Summarizing Agent Strategies," *Journal of Autonomous Agents and Multi-Agent Systems (AAMAS)*, vol. 33, pp. 628-644, July 2019.
- J. Wiens, S. Saria, M. Sendak, M. Ghassemi, V. Liu, **F. Doshi-Velez**, K. Jung, K. Heller, D. Kale, M. Saeed, P. Ossorio, S. Thadaney-Israni, and A. Goldenberg, "Do no harm: a roadmap for responsible machine learning for healthcare," *Nature Medicine*, vol. 25 (10), pp. 1337-1340, July 2019.
- *M. Masood*, and **F. Doshi-Velez**, "A Particle-Based Variational Approach to Bayesian Non-negative Matrix Factorization," *Journal of Machine Learning Research*, vol. 20 (90), pp. 1-56, May 2019.
- A. Fan, **F. Doshi-Velez**, and L. Miratrix, "Assessing topic model relevance: Evaluation and informative priors," *Statistical Analysis and Data Mining*, vol. 12, pp. 210-222, May 2019.
- *O. Gottesman*, F. Johansson, M. Komorowski, A. Faisal, D. Sontag, **F. Doshi-Velez**, and L. Celi. "Guidelines for reinforcement learning in healthcare," *Nature Medicine*, vol. 25, pp. 16–18, January 2019.
- *S. Parbhoo, O. Gottesman, AS Ross*, M. Komorowski, A. Faisal, I. Bon, V. Roth, **F. Doshi-Velez.** "Improving counterfactual reasoning with kernelised dynamic mixing models," *PLoS ONE*, vol. 13, no. 11, November 2018.
- **F. Doshi-Velez** and S. Williamson. "Restricted Indian Buffet Processes," *Statistics and Computing*, vol. 27(5), pp. 1205-1223, September 2017.
- T. Wang, C. Rudin, **F. Doshi-Velez**, Y. Liu, E. Klampfl, P. MacNeille. "A Bayesian Framework for Learning Rule Sets for Interpretable Classification," *Journal of Machine Learning Research*, vol. 18 (70), pp. 1-37, August 2017.
- M. Wu, M. Ghassemi, M. Feng, L. A. Celi, P. Szolovits, **F. Doshi-Velez**. "Understanding Vasopressor Intervention and Weaning: Risk Prediction in a Public Heterogeneous Clinical Time Series Database," *Journal of the American Medical Informatics Association*, Vol. 24(3), pp. 488-495, May 2017.
- J. Gafford, **F. Doshi-Velez**, R. Wood, C. Walsh. "Machine Learning Approaches to Environmental Disturbance Rejection in Multi-Axis Optoelectronic Force Sensors," *Sensors and Actuators A: Physical*, vol. 248, pp.78-87, September 2016.
- T. Lingren, P. Chen, J. Bochenek, **F. Doshi-Velez**, P. Manning-Courtney, J. Bickel, L. W. Welchons, J. Reinhold, N. Bing, Y. Ni, W. Barbaresi, F. Mentch, M. Basford, J. Denny, L. Vazquez, C. Perry, B. Namjou, H. Qiu, J. Connolly, D. Abrams, I. A. Holm, B. A. Cobb, N. Lingren, I. Solti, H. Hakonarson, I. S. Kohane, J. Harley, G. Savova. "Electronic Health Record Based Algorithm to Identify Patients with Autism Spectrum Disorder," *Public Library of Science ONE*, vol. 11(7), July 29, 2016.

---

[1] Italicized authors are graduate students and postdocs in my lab. Underlined authors are masters and undergraduate students in my lab. Authors that are collaborators at other institutions or visiting students are in plain font.

- H. M. Elibol, V. Nguyen, S. Linderman, M. Johnson, A. Hashmi, **F. Doshi-Velez**. "Cross-Corpora Unsupervised Learning of Trajectories in Autism Spectrum Disorders," *Journal of Machine Learning Research*, vol. 17(133) pp. 1-38, January 2016.
- **F. Doshi-Velez**, P. Avillach, N. Palmer, A. Bousvaro, Y. Ge, K. Fox, G. Steinberg, C. Spettell, I Juster, I. Kohane. "Prevalence of Inflammatory Bowel Disease Among Patients with Autism Spectrum Disorders," *Inflammatory Bowel Diseases*, vol. 21(10), pp. 2281-2288, October 2015.
- **F. Doshi-Velez**, F. Pfau, F.Wood, N. Roy. "Bayesian Nonparametric Methods for Partially-Observable Reinforcement Learning'" *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 37(2), pp. 394-407, February 2015.
- **F. Doshi-Velez**, Y. Ge, I. Kohane. "Comorbidity Clusters in Autism Spectrum Disorders: An Electronic Health Record Time-Series Analysis," *Pediatrics* vol. 133(1), e54-e63, January 2014.
- **F. Doshi-Velez**, W. Li, Y. Battat, J. Park, B. Charrow, D. Curtis, S. Hemachandra, B. Reimer, J. Velez, C. Walsh, D. Fredette, N. Roy, S. Teller. "Improving Safety and Operational Efficiency in Residential Care Settings with WiFi-based Localization," *Journal of the American Medical Directors Association* vol. 13(6), pp. 558-563, July 2012.
- **F. Doshi-Velez**, J. Pineau, N. Roy. "Reinforcement Learning with Limited Reinforcement: Using Bayes Risk for Active Learning in POMDPs," *Artificial Intelligence Journal*, vol. 187–188, pp. 115-132, August 2012.
- J. Joseph, **F. Doshi-Velez**, A. Huang, N. Roy. "A Bayesian Nonparametric Approach to Modeling Motion Patterns," *Autonomous Robots*, vol. 31(4), pp. 383-400, November 2011.
- **F. Doshi-Velez** and N. Roy. "Spoken Language Interaction with Model Uncertainty: An Adaptive Human-Robot Interaction System," *Connection Science*, vol. 20(4), pp. 299-318, December 2008.

Book Chapter

- **F. Doshi-Velez** and B. Kim. "Considerations for Evaluation and Generalization in Interpretable Machine Learning," in *Explainable and Interpretable Models in Computer Vision and Machine Learning*, 1st ed., H. Escalante, S. Escalera, I. Guyon, X. Baró, Y. Güçlütürk, U. Güçlü, and M. A. J. van Gerven, Eds. Springer International Publishing, September 2018.

Conference

- M. Jacobs, J. He, *M. Pradier*, B. Lam, A. Ahn, T. McCoy, R. Perlis and F. Doshi-Velez, and K. Gajos, "Designing AI for Trust in Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens," presented at the Conference on Conference on Human Factors in Computing Systems (CHI), 2021.
- *A. Ross*, N. Chen, E. Hang, E. Glassman and **F. Doshi-Velez**, "Evaluating the Interpretability of Generative Models by Interactive Reconstruction," presented at the Conference on Conference on Human Factors in Computing Systems (CHI), 2021. **(Selected for Honorable Mention, 5% of papers)**
- *M. Pradier*, J. Zazo, *S. Parbhoo*, R. Perlis, M. Zazzi and **F. Doshi-Velez**. "Preferential Mixture-of-Experts: Interpretable Models that Rely on Human Expertise as Much as Possible," presented at the American Medical Informatics Association (AMIA), 2021.
- W. Yang, L. Lorch, M. Graule, H. Lakkaraju, **F. Doshi-Velez**. Incorporating Interpretable Output Constraints in Bayesian Neural Networks. NeurIPS 2020. **(Selected as Spotlight, 3% of papers)**
- J. Du, J. Futoma, **F. Doshi-Velez**. Model-Based Reinforcement Learning for Semi-Markov Decision Processes with Neural ODEs. NeurIPS 2020.
- M. Lu, Z. Shahn, D. Sow, **F. Doshi-Velez**, and L. Lehman, "Is Deep Reinforcement Learning Ready for Practical Applications in Healthcare? A Sensitivity Analysis of Duel-DDQN for Sepsis Treatment," *AMIA*, pp 1-13, vol. 2, May 2020. **(Distinguished Paper Award)**
- N. Prasad, B. Engelhardt, and **F. Doshi-Velez**, "Defining Admissible Rewards for High-Confidence Policy Evaluation in Batch Reinforcement Learning," *ACM Conference on Health, Inference and Learning*, vol. 2, pp 1-9, April 2020.
- M. Wu, *S. Parbhoo, M. Hughes*, R. Kindle, L. Celi, M. Zazzi, R. Volker, and **F. Doshi-Velez**, "Regional Tree Regularization for Interpretability in Deep Neural Networks," *AAAI*, vol.3, pp 1-9, March 2020
- *Y. Yacoby, W. Pan*, and **F. Doshi-Velez**, "Characterizing and Avoiding Problematic Global Optima of Variational Autoencoders," *Advances in Approximate Bayesian Inference*, vol. 1, pp. 1-17, March 2020.
- *J. Futoma, M. Hughes*, and **F. Doshi-Velez**, "POPCORN: Partially Observed Prediction Constrained Reinforcement Learning," *AISTATS*, vol. 2, pp 1-18, March 2020.
- *Ross, W. Pan*, L. Celi, and **F. Doshi-Velez**, "Ensembles of Locally Independent Prediction Models," *AAAI*, vol. 3, pp 1-11, February 2020.

- J. Ren, R. Kunes, and **F. Doshi-Velez**, "Prediction Focused Topic Models via Feature Selection," *AISTATS*, vol. 2, pp 1-19, February 2020.
- *J. Futoma, M. Masood*, and **F. Doshi-Velez**, "Identifying Distinct, Effective Treatments for Acute Hypotension with SODA-RL: Safely Optimized Diverse Accurate Reinforcement Learning," *AMIA CRI*, vol.1, pp 1-24, January 2020.
- *B. Coker, M. Fernandez-Pradier,* and **F. Doshi-Velez**, "PoRB-Nets: Poisson Process Radial Basis Function Networks," *UAI*, pp 1-59, 2020.
- *S. Srinivasan,* and **F. Doshi-Velez**, "Interpretable Batch IRL to extract clinician goals in ICU Hypotension Management," *AMIA CRI*, vol. 1, pp 636-645, 2020.
- *O. Gottesman, J. Futoma*, Y. Liu, *S. Parbhoo,* LA. Celi, E. Brunskill, and **F. Doshi-Velez**, "Interpretable Off-Policy Evaluation in Reinforcement Learning by Highlighting Influential Transitions," presented at the International Conference on Machine Learning, 2020.
- *O. Gottesman*, Y. Liu, S. Sussex, E. Brunskill, and **F. Doshi-Velez.** "Combining parametric and nonparametric models for off-policy evaluation," presented at the International Conference on Machine Learning (ICML), 2019.
- *I. Lage*, E. Chen, J. He, M. Narayanan, B. Kim, S. Gershman, and **F. Doshi-Velez**. "Human Evaluation of Models Built for Interpretability," presented at the 7th AAAI Conference on Human Computation and Crowdsourcing (HCOMP), 2019. **(Selected as Best Paper Finalist)**
- *I. Lage*, D. Lifschitz, **F. Doshi-Velez**, and O. Amir. "Toward Robust Policy Summarization," presented at the International Joint Conference on Artificial Intelligence (IJCAI), 2019.
- *M. Masood* and **F. Doshi-Velez.** "Diversity-Inducing Policy Gradient: Using Maximum Mean Discrepancy to Find a Set of Diverse Policies," presented at the International Joint Conference on Artificial Intelligence (IJCAI), 2019.
- S. Srinivasan, D. Lee, and **F. Doshi-Velez**. "Truly Batch Apprenticeship Learning with Deep Successor Features," presented at the International Joint Conference on Artificial Intelligence (IJCAI), 2019.
- *I. Lage, A. Ross*, B. Kim, S. Gershman, and **F. Doshi-Velez.** "Human-in-the-Loop Interpretability Prior," presented at the Conference on Neural Information Processing Systems (NeurIPS), 2018. **(Selected for a Spotlight, 3% of papers)**
- Y. Liu, *O. Gottesman*, A. Raghu, M. Komorowski, A. Faisal, **F. Doshi-Velez**, and E. Brunskill. "Representation Balancing MDPs for Off-policy Policy Evaluation," presented at the Conference on Neural Information Processing Systems (NeurIPS), 2018.
- L. Jin, **F. Doshi-Velez**, T. Miller, W. Schuler, and L. Schwartz. "Depth-bounding is effective: Improvements and Evaluation of Unsupervised PCFG Induction," presented at the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018.
- X. Peng, Y. Ding, D. Wihl, *O. Gottesman*, M. Komorowski, L. Lehman, *A. Ross*, A. Faisal, and **F. Doshi-Velez**. "Improving Sepsis Treatment Strategies using Deep Reinforcement Learning and Mixture-of-Experts," presented at the American Medical Informatics Association (AMIA) Annual Symposium, 2018.
- S. Ghosh, *J. Yao*, and **F. Doshi-Velez**. "Structured Variational Learning of Bayesian Neural Networks with Horseshoe Priors," presented at the International Conference on Machine Learning (ICML), 2018.
- S. Depeweg, J. Hernandez-Lobato, **F. Doshi-Velez**, and S. Udluft. "Decomposition of Uncertainty in Bayesian Deep Learning for Efficient and Risk-sensitive Learning," presented at the International Conference on Machine Learning (ICML), 2018.
- *O. Gottesman, W. Pan*, and **F. Doshi-Velez**. "Weighted Tensor Decomposition for Learning Latent Variables with Partial Data," presented at the International Conference on Artificial Intelligence and Statistics (AISTATS), 2018.
- *M. Hughes*, J. Hope, L. Weiner, T. McCoy, R. Perlis, E. Sudderth, and **F. Doshi-Velez**. "Semi-Supervised Prediction-Constrained Topic Models," presented at the International Conference on Artificial Intelligence and Statistics (AISTATS), 2018.
- O. Amir, **F. Doshi-Velez**, and D. Sarne. "Agent Strategy Summarization," in Autonomous Agents and Multiagent Systems, Blue Sky Track, Stockholm, Sweden, 2018.
- **F. Doshi-Velez**, M. Kortz, R. Budish, C. Bavitz, S. Gershman, D. O'Brien, S. Schieber, J. Waldo, D. Weinberger, and A. Wood. "Accountability of AI Under the Law: The Role of Explanation," presented at the Privacy Law Scholars Conference, 2018.
- L. Jin, **F. Doshi-Velez**, T. Miller, W. Schuler, and L. Schwartz. "Unsupervised Grammar Induction with Depth-bounded PCFG," presented at the Association for Computational Linguistics, 2018.
- M. Wu, *M. Hughes, S. Parbhoo*, M. Zazzi, V. Roth, and **F. Doshi-Velez**. "Beyond Sparsity: Tree Regularization of Deep Models for Interpretability," presented at the Association for the Advancement of Artificial Intelligence (AAAI), 2018.
- *A. Ross*, and **F. Doshi-Velez**. "Improving the Adversarial Robustness and Interpretability of Deep Neural Networks by Penalizing their Certainty Sensitivity," presented at the Association for the Advancement of Artificial Intelligence (AAAI), 2018. **(Selected for an Oral)**

- T. Killian, S. Daulton, G. Konidaris, and **F. Doshi-Velez**. "Robust and Efficient Transfer Learning with Hidden-Parameter Markov Decision Processes," presented at the Advances in Neural Information Processing Systems (NIPS), 2017. **(Selected for an Oral, 1% of papers)**
- M. Ghassemi, M. Wu, *M. C. Hughes*, P. Szolovits, and **F. Doshi-Velez**. "Predicting Intervention Onset in the ICU with Switching State Space Models," presented at the American Medical Informatics Association (AMIA), 2017. **(Best Paper Nominee**)
- S. Parhboo, J. Bogojeska, M. Zazzi, V. Roth, and **F. Doshi-Velez**. "Combining Kernel and Model-Based Learning for HIV Therapy Selection," presented at the American Medical Informatics Association (AMIA) Joint Summits, 2017.
- A. Fan, **F. Doshi-Velez**, and L. Miratrix. "Prior Matters: Simple and General Methods for Evaluating and Improving Topic Quality in Topic Modeling," presented at the Text as Data conference, 2017.
- M. Glueck, M. Pakdaman Naeini, **F. Doshi-Velez**, F. Chevalier, A. Khan, D. Widgor, and M. Brudno. "Phenotype Comparison Visualizations for Disease Subtyping via Topic Models," presented at the Conference on Visual Analytics Science and Technology (VAST), 2017.
- *A. Ross, M. C. Hughes*, and **F. Doshi-Velez**, Finale. "Right for the Right Reasons: Training Differentiable Models by Constraining their Explanations," presented at the International Conference on Artificial Intelligence (IJCAI), 2017.
- S. Depeweg, M. Hernandez-Lobato, **F. Doshi-Velez**, and S. Udluft. "Learning and Policy Search in Stochastic Dynamical Systems with Bayesian Neural Networks," presented at the International Conference on Learning Representations (ICLR), 2017.
- T. Wang, C. Rudin, **F. Doshi-Velez**, Y. Liu, E. Klampfl, and P. MacNeille. "Bayesian Or's of And's for Interpretable Classification with Application to Context Aware Recommender Systems," presented at the International Conference on Data Mining (ICDM), 2016.
- X. Xia, P. Protopapas, and **F. Doshi-Velez**. "Cost-Sensitive Batch Mode Active Learning: Designing Astronomical Observation by Optimizing Telescope Time and Telescope Choice," presented at the International Conference on Data Mining (ICDM), 2016.
- Shain, W. Bryce, L. Jin, V. Krakovna, **F. Doshi-Velez**, T. Miller, W. Schuler, and L. Schwartz. "Memory-Bounded Left-Corner Unsupervised Grammar Induction on Child-Directed Input," presented at the International Conference on Computational Linguistics (COLING), 2016.
- **F. Doshi-Velez**, and G. Konidaris. "Hidden Parameter Markov Decision Processes: A Semiparametric Regression Approach for Discovering Latent Task Parametrizations," presented at the International Joint Conference on Artificial Intelligence (IJCAI), 2016.
- D. Tran, M. Kim, and **F. Doshi-Velez**. "Spectral M-estimation with Application to Hidden Markov Models: Supplementary Material," presented at the International Conference on Artificial Intelligence and Statistics (AISTATS), 2016
- B. Kim, J. A. Shah, and **F. Doshi-Velez**. "Mind the Gap: A Generative Approach to Interpretable Feature Selection and Extraction," presented at Advances in Neural Information Processing Systems (NIPS), 2015.
- **F. Doshi-Velez**, and Y. E. Marshall. "HackEbola with Data: On the Hackathon Format for Timely Data Analysis," presented at Humanitarian Technology: Science, Systems and Global Impact, 2015.
- **F. Doshi-Velez**, B. C. Wallace, and R. Adams. "Graph-Sparse LDA: A Topic Model with Structured Sparsity," presented at the Association for the Advancement of Artificial Intelligence (AAAI), 2015.
- M. Ghassemi, T. Naumann, **F. Doshi-Velez**, N. Brimmer, R. Joshi, A. Rumshisky, and P. Szolovits. "Unfolding Physiological State: Mortality Modeling in Intensive Care Units," presented at the International Conference on Knowledge Discovery and Data Mining (KDD), 2014.
- J. Joseph, **F. Doshi-Velez**, and N. Roy. "A Bayesian Nonparametric Approach to Modeling Battery Health," presented at the International Conference on Robotics and Automation (ICRA), 2012.
- **F. Doshi-Velez** and Z. Ghahramani. "A Comparison of Human and Agent Reinforcement Learning in Partially Observable Domains," presented at the 33rd Annual Meeting of the Cognitive Science Society (CogSci), 2011.
- **F. Doshi-Velez**, D. Wingate, J. B. Tenenbaum, and N. Roy. "Infinite Dynamic Bayesian Networks," presented at the International Conference on Machine Learning (ICML), 2011.
- Geramifard, **F. Doshi-Velez**, J. Redding, N. Roy, and J. P. How. "Online Discovery of Feature Dependencies" presented at the International Conference on Machine Learning (ICML), 2011.
- **F. Doshi-Velez**, D. Wingate, N. Roy, and J. Tenebaum. "Nonparametric Bayesian Policy Priors for Reinforcement Learning," presented at Advances in Neural Information Processing Systems (NIPS), 2010.
- J. Joseph, **F. Doshi-Velez**, and N. Roy. A Bayesian Nonparametric Approach to Modeling Mobility Patterns. presented at the Association for the Advancement of Artificial Intelligence (AAAI), 2010.
- **F. Doshi-Velez**. "The Infinite Partially Observable Markov Decision Process," presented at Advances in Neural Information Processing Systems (NIPS), 2009.

- **F. Doshi-Velez**, D. Knowles, S. Mohammed, and Z. Ghahramani. "Large Scale Nonparametric Bayesian Inference: Data Parallelisation in the Indian Buffet Process," presented at Advances in Neural Information Processing Systems (NIPS), 2009.
- **F. Doshi-Velez**, and Z. Ghahramani. "Correlated Non-Parametric Latent Feature Models," presented at Uncertainty in Artificial Intelligence (UAI), 2009.
- **F. Doshi-Velez**, and Z. Ghahramani. "Accelerated Gibbs Sampling for the Indian Buffet Process," presented at the International Conference on Machine Learning (ICML), 2009.
- **F. Doshi-Velez**, K. Miller, J. Van Gael, and Y. W. Teh. "Variational Inference for the Indian Buffet Process," International Conference on Artificial Intelligence and Statistics (AISTATS), 2009. (**Best Paper Nominee**)
- **F. Doshi**, J. Pineau, and N. Roy, Nicholas. "Reinforcement Learning with Limited Reinforcement: Using Bayes-Risk for Active Learning in POMDPS," presented at the International Conference on Machine Learning (ICML), 2008.
- **F. Doshi**, and N. Roy. "The Permutable POMDP: Fast Solutions to POMDPs for Preference Elicitation," presented at Autonomous Agents and Multi-Agent Systems (AAMAS), 2008. (**Best Paper Nominee**)
- **F. Doshi**, E. Brunskill, A. Shkolnik, T. Kollar, K. Rohanimanesh, R. Tedrake, and N. Roy. "Collision Detection in Legged Locomotion using Supervised Learning," presented at Intelligent Robots and Systems (IROS), 2007.
- **F. Doshi**, and N. Roy. "Efficient Model Learning for Dialog Management," presented at Human Robot Interaction (HRI), 2007.

Referred Workshop and Symposia

- *S. Parbhoo*, *O. Gottesman* and F. Doshi-Velez. "Shaping Control Variates for Off-Policy Evaluation," presented at the Conference on Neural Information Processing Systems (NeurIPS) Workshop on Offline Reinforcement Learning, 2020.
- K.Zhang, H. Wang, J. Du, B. Chu, R. Kindle, L. Celi, and F. Doshi-Velez. "Identifying Decision Points for Safe and Interpretable Reinforcement Learning in Hypotension Treatment," presented at the Conference on Neural Information Processing Systems (NeurIPS) Workshop on Machine Learning for Health, 2020.
- K. Zhang, *O. Gottesman*, and **F. Doshi-Velez.** "A Bayesian Approach to Learning Bandit Structure in Markov Decision Processes," presented at the Conference on Neural Information Processing Systems (NeurIPS) Workshop on Real World Reinforcement Learning, 2020.
- J. Antoran, *J. Yao*, *W. Pan*, **F. Doshi-Velez**, and J. Hernandez-Lobato, "ICML Workshop on Uncertainty in Deep Learning," *ICML Workshop on Uncertainty in Deep Learning*, pp 1-11, 2020.
- *Y. Yacoby*, *W. Pan*, and **F. Doshi-Velez**, "Failures of Variational Autoencoders and their Effects on Downstream Tasks," presented at the International Conference on Machine Learning (ICML) Workshop on Uncertainty in Deep Learning, 2020.
- T. Guenais, D. Vamvourellis, *Y. Yacoby*, **F. Doshi-Velez**, and *W. Pan*, "BaCOUn: Bayesian Classifers with Out-of-Distribution Uncertainty," presented at the International Conference on Machine Learning (ICML) Workshop on Uncertainty in Deep Learning, 2020.
- S. Thakur, C. Lorsung, *Y. Yacoby*, **F. Doshi-Velez**, and *W. Pan*, "Learned Uncertainty-Aware (LUNA) Bases for Bayesian Regression using Multi-Headed Auxiliary Networks," presented at the International Conference on Machine Learning (ICML) Workshop on Uncertainty in Deep Learning, 2020.
- Y. Nair, and F. Doshi-Velez, "PAC Imitation and Model-based Batch Learning of Contextual MDPs," presented at the International Conference on Machine Learning (ICML) Workshop on Inductive Biases, Invariances and Generalization in RL, 2020.
- Y. Nair, and **F. Doshi-Velez**, "PAC Imitation and Model-based Batch Learning of Contextual MDPs," presented at the International Conference on Machine Learning (ICML) Workshop on Theoretical Foundations of Reinforcement Learning, 2020.
- *J. Yao*, E. Brunskill, *W. Pan*, S. Murphy, and **F. Doshi-Velez**, "Power-Constrained Bandits" presented at the International Conference on Machine Learning (ICML) Workshop on Theoretical Foundations of Reinforcement Learning, 2020.
- HC. Ou, K. Wang, **F. Doshi-Velez,** and M. Tambe, "Active Screening on Recurrent Diseases Contact Networks with Uncertainty: A Reinforcement Learning Approach," presented at the Conference on International Conference on Autonomous Agents and Multiagent Systems (AAMAS) Workshop on Multi-Agent Based Simulation, 2020.
- *I. Lage*, and **F. Doshi-Velez**, "Human-in-the-Loop Learning of Interpretable and Intuitive Representations," presented at the International Conference on Machine Learning (ICML) Workshop on Human Interpretability in Machine Learning, 2020.

- M. Downs, J. Chu, *Y. Yacoby*, **F. Doshi-Velez**, and *W. Pan*, "CRUDS: Counterfactual Recourse Using Disentangled Subspaces," presented at the International Conference on Machine Learning (ICML) Workshop on Human Interpretability in Machine Learning, 2020.
- N. Prasad, B. Engelhardt, and **F. Doshi-Velez**, "Defining Admissible Rewards for High Confidence Policy Evaluation," presented at the Conference on Neural Information Processing Systems (NeurIPS) Workshop on Safety and Robustness in Decision-Making, 2019.
- *M. Pradier, W. Pan, J. Yao,* S. Ghosh, and **F. Doshi-Velez**, "Projected BNNs: Avoiding Weight-space Pathologies by Learning Latent Representations of Neural Network Weights," presented at the Conference on Asian Conference on Machine Learning (ACML) Workshop on Weakly Supervised Learning Workshop, 2019.
- *A. Ross*, J. Du, Y. Shavit, and **F. Doshi-Velez**, and "Controlled Direct Effect Priors for Bayesian Neural Networks," presented at the Conference on Neural Information Processing Systems (NeurIPS) Workshop on Bayesian Deep Learning, 2019.
- J. Ren, R. Kunes and **F. Doshi-Velez**, "Prediction Focused Topic Models Via Vocab Filtering," *NeurIPS Workshop on Human-Centric ML*, vol. 1, pp. 1-12, November 2019.
- M. Jacobs, R. Perlis, *M. Pradier,* **F. Doshi-Velez**, E. Mynatt, and K. Gajos "Integrating ai recommendations into the pharmacologic management of major depressive disorder," presented at the Conference on Computer Supported Cooperative Work (CSCW) Workshop on Identifying Challenges and Opportunities in Human–AI Collaboration in Healthcare, 2019.
- *M. Pradier, M. Hughes*, and **F. Doshi-Velez**, "Challenges in Computing and Optimizing Upper Bounds of Marginal Likelihood based on Chi-Square Divergences," presented at the Conference on Advances in Approximate Bayesian Inference (AABI), 2019.
- J. Ren, R. Kunes, and **F. Doshi-Velez**, "Prediction Focused Topic Models for Electronic Health Records," presented at the Conference on Neural Information Processing Systems (NeurIPS) Workshop on Machine Learning for Health, 2019.
- D. Vaughan, *W. Pan, Y. Yacoby*, E. Seidler, A. Leung, **F. Doshi-Velez**, and D. Sakkas. "The Application of Machine Learning Methods to Evaluate Predictors for Live Birth in Programmed Thaw Cycles," presented at the American Society for Reproductive Medicine Scientific Congress & Expo (ASRM), 2019.
- W. Yang, L. Lorch, M. Graule, S. Srinivasan, A. Suresh, *J. Yao, M. Pradier*, and **F. Doshi-Velez**. "Output-Constrained Bayesian Neural Network," presented at the International Conference on Machine Learning (ICML) Workshop on Understanding and Improving Generalization in Deep Learning, 2019.
- W. Yang, L. Lorch, M. Graule, S. Srinivasan, A. Suresh, *J. Yao, M. Pradier,* and **F. Doshi-Velez**. "Output-Constrained Bayesian Neural Networks," presented at the International Conference on Machine Learning (ICML) Workshop on Uncertainty & Robustness in Deep Learning, 2019.
- *Y. Yacoby, W. Pan*, and **F. Doshi-Velez**. "Mitigating Model Non-Identifiability in BNN with Latent Variables," presented at the International Conference on Machine Learning (ICML) Workshop on Uncertainty & Robustness in Deep Learning, 2019.
- *J. Yao, W. Pan*, S. Ghosh, and **F. Doshi- Velez**. "Quality of Uncertainty Quantification for Bayesian Neural Network Inference," presented at the International Conference on Machine Learning (ICML) Workshop on Uncertainty & Robustness in Deep Learning, 2019.
- *B. Coker, M. Pradier,* and **F. Doshi-Velez**. "Poisson Process Bayesian Neural Networks," presented at the International Conference on Bayesian Nonparametrics, (BNP), 2019.
- *I. Lage*, D. Lifschitz, **F. Doshi-Velez**, and O. Amir. "Toward Robust Summarization of Agent Policies," presented at the International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2019.
- D. Lifschitz, *I. Lage*, **F. Doshi-Velez,** and O. Amir. "Exploring Computational User Models for Agent Policy Summarization," presented at the International Joint Conference on Artificial Intelligence: Workshop on Explainable Artificial Intelligence (IJCAI), 2019.
- Z. Juozapaitis, A. Koul, A. Fern, M. Erwig, and **F. Doshi-Velez**. "Explainable Reinforcement Learning via Reward Decomposition," presented at the International Joint Conference on Artificial Intelligence: Workshop on Explainable Artificial Intelligence (IJCAI), 2019.
- *I. Lage*, E. Chen, J. He, M. Narayanan, S. Gershman, B. Kim and **F. Doshi-Velez.** "An Evaluation of the Human-Interpretability of Explanation," presented at the Conference on Neural Information Processing Systems (NeurIPS) Workshop on Correcting and Critiquing Trends in Machine Learning, 2018.
- *M. Fernandez-Pradier, W. Pan, J. Yao*, S. Ghosh, and **F. Doshi-Velez.** "Projected BNNs: Avoiding weight-space pathologies by projecting neural network weights," presented at the Conference on Neural Information Processing Systems (NeurIPS) Workshop on Bayesian Deep Learning, 2018.
- *M. Fernandez-Pradier, W. Pan, J. Yao*, R. Singh, and **F. Doshi-Velez.** "Hierarchical Stick-breaking Feature Paintbox," presented at the Conference on Neural Information Processing Systms (NeurIPS) Workshop on All of Bayesian Nonparametrics, 2018.

- *J. Futoma, M. Hughes,* and **F. Doshi-Velez.** "Prediction-Constrained POMDPs," presented at Conference on Neural Information Processing Systems (NeurIPS) Workshop on Reinforcement Learning under Partial Observability, 2018.
- *J. Yao*, T. Killian, G. Konidaris, and **F. Doshi-Velez**. "Direct Policy Transfer via Hidden Parameter Markov Decision Processes," presented at the International Conference on Machine Learning (ICML) Workshop on Lifelong Learning, 2018. (Selected by Lincoln Labs as their "Paper of the Year") HH
- A. Raghu, *O. Gottesman,* Y. Liu, M. Komorowski, A. Faisal, **F. Doshi-Velez,** and E. Brunskill. "Behaviour Policy Estimation in Off-Policy Policy Evaluation: Calibration Matters," presented at the International Conference on Machine Learning (ICML) Workshop on CausalML, 2018.
- S. Sussex, *O. Gottesman*, Y. Liu, S. Murphy, E. Brunskill, and **F. Doshi-Velez**. "Stitched Trajectories for Off-Policy Learning," presented at the International Conference on Machine Learning (ICML) Workshop on CausalML, 2018.
- Y. Liu, *O. Gottesman*, A. Raghu, M. Komorowski, A. Faisal, **F. Doshi-Velez**. "Representation Balancing MDPs for Off-Policy Policy Evaluation," presented at the International Conference on Machine Learning (ICML) Workshop on CausalML, 2018.
- *O. Gottesman* and **F. Doshi-Velez**, "Regularizing tensor decomposition methods by optimizing pseudo-data," presented at the International Conference on Machine Learning (ICML) Modern Trends in Nonconvex Optimization for Machine Learning Workshop, 2018.
- *M. Masood* and **F. Doshi-Velez**, "Diversity-Inducing Policy Gradient: Using MMD to find a set of policies that are diverse in terms of stete-visitation," presented at the International Conference on Machine Learning (ICML) Exploration in Reinforcement Learning Workshop, 2018.
- *A. Ross, W. Pan,* and **F. Doshi-Velez**, "Learning Qualitatively Diverse and Interpretable Rules for Classification," presented at the ICML Workshop on Human Interpretability in Machine Learning, 2018.
- **F. Doshi-Velez**, *M. Hughes,* and R. Perlis. "Predicting Depression Treatments from EHR," presented at the American College of Neuropsychopharmacology Annual Meeting, Symposium on Pragmatic Academic Research in Psychiatry, 2017.
- *A. Ross, E. Lage,* and **F. Doshi-Velez**. "The Neural LASSO: Local Linear Sparsity for Interpretable Explanations," presented at the Neural Information Processing Systems (NIPS) Workshop on Transparent and Interpretable Machine Learning in Safety Critical Environments, 2017.
- M. Wu, *M. Hughes*, S. Parbhoo, M. Zazzi, V. Roth, and **F. Doshi-Velez**. "Beyond Sparsity: Tree Regularization of Deep Models for Interpretability," presented at the Neural Information Processing Systems (NIPS) Workshop on Transparent and Interpretable Machine Learning in Safety Critical Environments, 2017.
- S. Parbhoo, V. Roth, and **F. Doshi-Velez**. "Counterfactual Reasoning with Dynamic Switching Models for HIV Therapy Selection," presented at the Neural Information Processing Systems (NIPS) Workshop on Machine Learning for Healthcare, 2017.
- *M. C. Hughes*, G. Hope, L. Weiner, T. H. McCoy, Roy. H. Perlis, E. B. Sudderth, and **F. Doshi-Velez**. "Prediction-Constrained Topic Models for Antidepressant Recommendation," presented at the Neural Information Processing Systems (NIPS) Workshop on Machine Learning for Healthcare, 2017.
- S. Ghosh, and **F. Doshi-Velez**. "Model Selection in Bayesian Neural Networks via Horseshoe Priors," presented at the Neural Information Processing Systems (NIPS) Workshop on Bayesian Deep Learning, 2017.
- R. Singh, J. Ling, and **F. Doshi-Velez**. "Structured Variational Autoencoders for the Beta-Bernoulli Process," presented at the Neural Information Processing Systems (NIPS) Workshop on Advances in Approximate Bayesian Inference, 2017.
- S. Tan, **F. Doshi-Velez**, J. Quiroz, and E. Glassman. "Clustering LaTeX Solutions to Machine Learning Assignments for Rapid Assessment," presented at the Advancing Education with Data Knowledge Discovery and Data Mining (KDD) Workshop, 2017.
- S. Depeweg, M. Hernandez-Lobato, **F. Doshi-Velez**, and S. Udluft. "Uncertainty Decomposition in Bayesian Neural Networks with Latent Variables," presented at the International Conference on Machine Learning (ICML) Workshop, 2017.
- S. Parhboo, J. Bogojeska, M. Zazzi, V. Roth, and **F. Doshi-Velez**. "Combining Kernel and Model-Based Learning for HIV Therapy Selection," presented at the Neural Information Processing Systems (NIPS) Workshop for Machine Learning and Healthcare, 2016. (**Best Paper**)
- T. Killian, G. Konidaris, and **F. Doshi-Velez**. "Transfer Learning Across Patient Variations with Hidden Parameter Markov Decision Processes," presented at the Neural Information Processing Systems (NIPS) Workshop for Machine Learning and Healthcare, 2016.
- T. Killian, G. Konidaris, and **F. Doshi-Velez**. "Robust and Efficient Transfer Learning with Hidden Parameter Markov Decision Processes," presented at the AAAI Student Abstract and Poster Program, 2017.

- *M. C. Hughes*, H. M. Elibol, T. McCoy, R. Perlis, and **F. Doshi-Velez**. "Supervised Topic Models for Clinical Interpretability," presented at the Neural Information Processing Systems (NIPS) Workshop for Machine Learning and Healthcare, 2016.
- *A. Masood*, and **F. Doshi-Velez**. "Robust Posterior Exploration in NMF," presented at the International Conference on Machine Learning (ICML) Workshop on Geometry in Machine Learning, 2016.
- *V. Krakovna*, and **F. Doshi-Velez**. "Increasing the Interpretability of Recurrent Neural Networks Using Hidden Markov Models," presented at the International Conference on Machine Learning (ICML) Workshop on Interpretable Machine Learning, 2016.
- G. Konidaris, and **F. Doshi-Velez**. "Hidden Parameter Markov Decision Processes: An Emerging Paradigm for Modeling Families of Related Tasks," presented at the AAAI 2014 Fall Symposium on Knowledge, Skill, and Behavior Transfer in Autonomous Robots, 2014.
- **F. Doshi-Velez**, and G. Konidaris. "Transfer Learning by Discovering Latent Task Parameterizations," presented at the Neural Information Processing Systems (NIPS) Workshop: Bayesian Nonparametric Models for Reliable Planning and Decision-Making Under Uncertainty, 2012.
- **F. Doshi-Velez**, and N. Roy. "An Analysis of Activity Changes in MS Patients: A Case Study in the Use of Bayesian Nonparametrics," presented at the Neural Information Processing Systems (NIPS) Workshop: Bayesian Nonparametrics, Hope or Hype? 2011.
- **F. Doshi-Velez**. "Bayesian Nonparametric Approaches to Reinforcement Learning in Partially Observable Domains," presented at the Association for the Advancement of Artificial Intelligence Doctoral Consortium, 2010.
- **F. Doshi**, and J. Van Gael. "Nonparametric Bayesian Methods for Finding Software Bugs," presented at the Centre for Research in Statistical Methodology (CRiSM) Workshop on High Dimensional Data, 2008.
- **F. Doshi**, and N. Roy. "Reinforcement Learning with Limited Reinforcement: Using Bayes Risk for Active Learning in POMDPs," presented at the International Symposium on Artificial Intelligence and Mathematics (ISAIM), 2008
- **F. Doshi**, and N. Roy. "Learning User Models with Limited Reinforcement: An Adaptive Human-Robot Interaction System," presented at the Symposium on Language and Robotics (LANGRO), 2007.
- **F. Doshi**, and N. Roy. "Efficient Model Learning for Dialog Management," presented at the Association for the Advancement of Artificial Intelligence Spring Symposium, 2007.
- **F. Doshi**, and N. Roy. "Model Learning for Dialog Management," presented at the Advances in Neural Information Processing Systems (NIPS) Workshop on Reinforcement Learning, 2006.

Theses

- Bayesian Nonparametric Approaches for Reinforcement Learning in Partially Observable Domains. PhD Thesis, MIT, 2012.
- The Indian Buffet Process: Scalable Inference and Extensions. Masters Thesis, Cambridge, 2009.
- Efficient Model Learning for Dialog Management. Masters Thesis, MIT, 2007

Patents

- Assessing Compressed-Database Raw Size, #20130212075. Lyric Pankaj Doshi and **Finale Doshi-Velez**. Issued August 2013.

Technical Reports and Other Publications

- E. Bertino**, F. Doshi-Velez**, M.Gina, D. Lopresti, D. Parkes, "Artificial Intelligence and Cooperation," arXiv preprint arXiv:2012.06034
- **F. Doshi-Velez** and B. Kim. "ML Techniques for Accountability," 1st ed., in Al Magazine Publishing, 2020.O.
- *O. Gottesman*, F. Johansson, J. Meier, J. Dent, D. H. Lee, S. Srinivasan, L. Zhang, Y. Ding, D. Wihl, X. Peng, *J. Yao*, *I. Lage*, C. Mosch, L. H. Lehman, M. Komorowski, A. Faisal, L. A. Celi, D. Sontag, **F. Doshi-Velez**. "Evaluating Reinforcement Learning Algorithms in Observational Health Settings," arXiv preprint arXiv:1805.12298
- *M. A. Masood*, and **F. Doshi-Velez**. "Rapid Posterior Exploration in Bayesian Non-negative Matrix Factorization," arXiv preprint arXiv:1610.08928.
- *M. A. Masood*, *W. Pan*, and **F. Doshi-Velez**. "An Empirical Comparison of Sampling Quality Metrics: A Case Study for Bayesian Nonnegative Matrix Factorization," arXiv preprint arXiv:1606.06250.
- *W. Pan*, and **F. Doshi-Velez**. "A Characterization of the Non-Uniqueness of Nonnegative Matrix Factorizations," arXiv preprint arXiv:1604.00653.

Professional Service, Teaching, and Leadership

- **Conferences Organized:**
  - Machine Learning for Healthcare 2016, 2017, 2018, 2019, 2020 (co-organizer); 2021 (advisory)
  - ICML Workshop Chair, 2018
- **Workshops Organized:**
  - NeurIPS Workshop 2020 on I Can't Believe It's Not Better! Bridging the gap between theory and empiricism in probabilistic machine learning
  - AAAI Fall Symposium 2014 on Knowledge, Skill, and Behavior Transfer in Autonomous Robots
  - NIPS Workshop 2013 on Machine Learning for Clinical Data Analysis
  - NIPS Workshop 2012 on Bayesian Nonparametric Models
  - ICML Workshop 2011 on Decision-Making with Uncertain Models.
  - AAAI Spring Symposium 2011 on Computational Physiology.
  - Workshop for Women in Machine Learning: Executive Board Member, 2015-present. Executive Board President 2015-2015, Co-founder and interim Co-chair of the Executive Board, Organizer 2009, Fundraiser, 2008. Organizer and fundraiser duties involved raising $25K from industry sponsors, program chairing, managing publicity; Executive Board duties included co-writing (multiple, successful) NSF grant for multi-year funding and organizing multi-year financial and information technology infrastructure.
  - Interdisciplinary Graduate Conference, Cambridge, 2009. (Program chair and organizer)
- **Reviewer** for the journals: Journal of Machine Learning Research, Autonomous Robots, Autonomous Agents; the conferences: AAAI, RSS (occasionally), IROS (occasionally), NIPS, AISTATS, ICML, JMLR.
- **Senior Program Committee/Area Chair/Senior Area Chair** for: IJCAI 2016; NIPS 2016, 2017, 2018, 2019, 2020, 2021; AAAI 2017, 2018, 2019, 2020; AISTATS 2018; 2019; 2020, 2021
- **Other Service:** NSF Small Panel (2018), CRA Computing Community Consortium (CCC) visioning meeting (2019), AI100 Study Panelist, ad-hoc reviewing
- **Departmental Service:** Graduate Admissions Diversity Committee (2016), Junior Faculty Search Committee (2015-2016, 2017-2018,2019-2020), Curriculum Committee (2020-2021); provided input/assisted with the organization of: SEAS visit day (2017), ML Exploratory Committee (2017-2018), Data Science Standing Committee (2017-2018; 2018-2019); Faculty founder and advisor of InTouch, Grad Student Peer Support Group at SEAS (2017-). ML Senior Faculty Search Committee (2018-2019, advisory)
- **Departmental Teaching:** CS282 Reinforcement Learning (Spring 2015), CS281 Graduate Machine Learning (Fall 2015), CS181 Undergraduate Machine Learning (Spring 2016, Spring 2018, Spring 2019, Spring 2020 [Commendation for Extraordinary Teaching 281 of over 2500], Spring 2021), CS282 Bayesian Nonparametrics (Spring 2017), CS282 Reinforcement Learning for Healthcare (Fall 2017), CS282 Deep Bayesian Models (Fall 2018), CS282 Batch Reinforcement Learning (Fall 2019), CS282 Loss-Constrained Models (Fall 2021)
- **Students and Postdocs supervised:**
  - **Undergraduate:** Joy Ming (2015, thesis), Borui Wang (2015), Melih Elibol (2015, thesis), Mike Wu (2016), Angela Fan (2016, thesis), Andre Nguyen (2016, thesis), Vincent Nguyen (2016, thesis), Amna Hashmi (2016, thesis), Sindy Tan (2017), Jesse Zhang (2017), George Qi (2017), Cecilia Zhou (2017), Jimmy Lin (2017, thesis), Madhu Vijay (2017, thesis), Frederick Widjaja (2017, thesis), Menaka Narayanan (2018), Jeffrey He (2018-2020, commendation for extraordinary teaching), Emily Chen (2018), Ben Barret (2019), William Deuschle (2019, thesis), Scott Sussex (2019), Lars Lorch (2019), Jason Ren (2019-2020, thesis), Michael Wornow (2019), Yash Nair (2019-), Wanqian Yang (2019-2020, thesis won Hoopes), Nari Johnson (2019-), Kristine Zhang (2019-, thesis), Sanjana Narayanan (2019-), Brian Tobin (2020-), Simon Shen (2019-2020), Jason Ma (2019-2020; thesis won Hoopes), Chloe Loughridge (2020-), Andrew Courtney (2020-2020), Marcus Trenfield (2020-2020), Frank Zhu (2020-), Catherine Zeng (2021-), Alex Rojas (2021-), Zad Chen (2021-), Vicki Xu (2021-).
  - **Master's:** Xide Xia (2015-2016), Taylor Killian (2016-2017), Weiwei Pan (2016-2017), Andrew Ross (2016-2018) , Srivatsan Srinivasan (2017-2019), Donghun Lee (2017-2019), Henry Wang (2018-2020), Jianzhun Du (2019-2020), Brian Chu (2020-2021), Esther Brown (2021-)
  - **PhD:** Arjumand Masood (2015-2019; Boston Consulting Group), Omer Gottesman (2016-2020; Postdoc at Brown University with George Konidaris and Michael Littman), Isaac Lage (2017-), Jiayu Yao (2017-), Yaniv Yacoby (2018-), Andrew Ross (2018-2021; Postdoc at Columbia), Abhishek Sharma (2020-), Eura Shin (2020-), Sarah Rathnam (2020-), Anna Li (2020-)
  - **Postdocs/Extended Visitors:** Marton Havasi (2021-), Sonali Parbhoo (2019-2021; Faculty at Imperial), Joe Futoma (2018-2020, Apple mHealth Research), Melanie Pradier (2017-2020, MSR Cambridge UK), Weiwei Pan (2018-), Michael C. Hughes (2016-2018, Assistant Professor at Tufts), Mahdi Pakdaman (2017-2018, Applied Scientist, Microsoft), Juan Quiroz (2017, Research Fellow, Australian Institute of Health Innovation).

- **Educational Outreach**
  - **International**: Machine Learning Summer School, Arequipa 2016, Duke-Kunshan 2017, Buenos Aires 2018; Indaba Deep Learning Nairobi 2019; Indaba Mentor 2020, 2021
  - **Executive Education:** Harvard RCC (2016, 2017, 2018, 2019), Digital Transformation in Government (2018, 2019), Artificial Intelligence: Exploring Policy, Technology, and Governance (2019, 2021)
  - **K12 Education:** Statistics Outreach at Cambridgeport Elementary School (May 2017, January 2018, May 2019)

Invited Seminars, Panels, and Guest Lectures

- Invited Speaker, ICML XAI Workshop, July 2021
- Invited Speaker, CBL Alumni Series Cambridge University, June 2021
- Invited Speaker, Maggie L. Walker Governor's School for Government and International Studies Mu Alpha Theta, April 2021
- Invited Panelist, Robust ML ICLR Workshop, April 2021
- Invited Speaker, Responsible AI ICLR Workshop, April 2021
- Invited Keynote, EPFL Applied Machine Learning Day Clinical ML Track, March 2021
- Guest Lecture, Stanford Reinforcement Learning Course, March 2021
- Invited Speaker, Pitt-CMU Machine Learning in Medicine Seminar, Feburary 2021
- Invited Speaker and Panelist, NeurIPS Human and Machine in-the-Loop Evaluation and Learning Strategies Workshop, December 2020
- Invited Speaker and Panelist, NeurIPS Offline RL Workshop, December 2020
- Mentor, Women in Machine Learning Networking Session, December 2020
- Invited Panelist, NeurIPS New in ML Workshop, December 2020
- Invited Speaker, NUS-NUH-MIT Critical Data Datathon, December 2020
- Invited Speaker, Stanford Biostat Workshop, November 2020
- Guest Lecture, Harvard Kennedy School DPI-662: Digital Government (Instructor: David Eaves), November 2020
- Invited Speaker, Oxford Women in CS Distinguished Speaker Series, November 2020
- Invited Speaker, Duke Biostatistics and Bioinformatics Seminar, October 2020
- Keynote, Visual Data Science at IEEE VIS, October 2020
- Invited Speaker, DARPA XAI Meeting, October 2020
- Invited Panelist, ICML July 2020
- Invited Speaker and Panelist, ICML Uncertainty in Deep Learning Workshop, July 2020
- Invited Speaker and Panelist, ICML WHI Workshop, July 2020
- Invited Panelist, RL in Real-life Workshop, June 2020
- Invited Speaker, Math Invitational for Girls, June 2020
- Invited Panelist, Science meets Engineering of Deep Learning, NeurIPS Workshops December 2020
- Invited Speaker, Human-Centric Machine Learning, NeurIPS Workshops December 2019
- Invited Speaker, Safety and Robustness in Decision-Making, NeuriPS Workshops December 2020
- Keynote, PLAGH-MIT Critical Care Datathon Beijing November, 2019
- Guest Lecture, Harvard Law School National Security Law, October 2019 (Head Instructor: Jim Baker)
- Keynote, Indaba Deep Learning, August 2019
- Guest Speaker, Michigan RLD3 Group, August 2019
- Invited Speaker, Harvard Standing Committee on Women, April 2019
- Invited Speaker, Toronto Fields-Vector Lecture Series, March 2019
- Invited Speaker, BI Harvard Medical Faculty Physicians Seminar, January 2019
- Invited Speaker, BWH Section of Clinical Science Seminar, January 2019
- Invited Speaker and Panelist, NIPS Critiquing and Correcting Trends in Machine Learning Workshop, December 2018
- Invited Speaker and Panelist, NIPS Machine Learning for Health Workshop, December 2018
- Guest Lecture, MIT MAS.S10 AI and Equality, November 2018 (Head Instructor: Kim Benard)
- Guest Lecture, Harvard Kennedy School DPI-662: Digital Government (Head Instructor: David Eaves)
- Invited Panelist, Biomedicine Symposium, Georgetown University, October 2018
- Invited Speaker, Engineering Seminar, George Washington University, October 2018
- Invited Panelist, Oregon State University 150 Celebration, October 2018
- Invited Speaker, EECS Colloquium; AI Group at Oregon State University, October 2018
- Guest Lecture, MIT HST 953 Collaborative Data Science in Medicine, Novemeber 2018 (Head Instructor: Leo Celi)
- Invited Tutorial on Interpretable Machine Learning, Harvard Data Science Initiative, October 2018

- Guest Lecture, Harvard Law School National Security Law, October 2018 (Head Instructor: Jim Baker)
- Invited Speaker, Amazon Research Cambridge UK, July 2018
- Invited Speaker, Microsoft Research Cambridge UK AI Frontiers Series, July 2018
- Invited Speaker, Imperial College London, July 2018
- Guest Speaker, MAVERIC Research Group at the VA, June 2018
- Invited Speaker, MIT MEDRC Workshop, May 2018
- Invited Moderator/Panel Organizer, New England Computational Health Summit, April 2018
- Invited Speaker, Commonwealth High School, April 2018
- Invited Speaker, Harvard Statistics Colloquium, April 2018
- Guest Speaker, Oracle Labs, March 2018
- Invited Panelist, SXSW Panel on Machine Learning: Hope or Hype?, March 2018
- Invited Speaker, UT Austin Statistics and Data Science Seminar Series, March 2018
- Keynote Panelist, IEN AI in Healthcare Summit, January 2018
- Keynote Speaker, Harvard IACS ComputeFest Digital Doctor: Healthcare in an Age of AI and Big Data, January 2018
- Invited Speaker, NIPS Workshop on Disentangled Representations, December 2017
- Invited Speaker and Panelist, NIPS Workshop on Bayesian Deep Learning, December 2017
- Invited Speaker and Panelist, NIPS Workshop on Transparent and Interpretable Machine Learning for Safety Critical Domains, December 2017
- Guest Speaker, Machine Learning Group at UC Riverside, December 2017
- Invited Speaker, Lincoln Laboratories Machine Learning Seminar, November 2017
- Invited Speaker, Harvard IACS Seminar Series, October 2017
- Invited Speaker, TEDxBoston, October 2017
- Invited Panelist, Google PAIR Symposium, September 2017
- Invited Speaker and Panelist, FDA Adept4 Workshop, September 2017
- Invited Speaker, UMD NLP Seminar Series, September 2017
- Invited Speaker, Cornell Medical Machine Learning Seminar, August 2017
- Invited Speaker, Google Machine Learning Seminar Series, April 2017
- Invited Speaker, Harvard Biostats and Public Health Seminar Series, March 2017
- Invited Speaker, AAAI Workshop on AI for Social Good, February 2017
- Invited Speaker, USC Machine Learning Seminar Series, February 2017
- Invited Speaker, Women in Data Science Conference, Stanford, February 2017
- Invited Speaker, NIPS Workshop on Deep Bayesian Learning, December 2016
- Keynote Speaker and Panelist, NIPS Workshop on Interpretable Machine Learning, December 2016
- Seminar Speaker, Harvard Statistics Seminar, October 2016
- Invited Speaker, Oracle, October 2016
- Invited Speaker, Columbia Department of Biomedical Informatics, July 2016
- Invited Speaker, Stanford-Whitehouse Future of AI Symposium, July 2016
- Invited Speaker, IJCAI Early Career Spotlight, July 2016
- Invited Speaker, New England Machine Learning Day, May 2016
- Invited Speaker, Microsoft Research New England Bioinformatics Seminar, May 2016
- Invited Speaker, TEDxHarvard, September 2016
- Invited Panelist, GET Conference, June 2016
- Invited Speaker, NIPS Workshop on Machine Learning for Healthcare, December 2015
- Invited Speaker, Carlos III Madrid, Novemeber 2015
- Invited Speaker, University of Massachusetts Medical School, October 2015
- Invited Speaker, Harvard Institute of Quantitative Social Science Seminar: April 2015, March 2016
- Guest Lecture, November 2015, CS182 AI
- Invited Speaker, JSM, June 2014
- Invited Speaker, February 2014, Duke University.
- Invited Speaker, December 2013, NIPS Workshop on Causal Reason. Tutorial of POMDPs for Causal Reasoning.
- Panelist, October 2013, Health 2.0 Conference. Machine Learning for Medical Data Panel.
- Guest Lecturer, October 2013, Harvard University, CS281: Advanced Machine Learning (advanced graduate course). Prepared and presented a lecture on Monte Carlo techniques.
- Speaker, July 2013, BIGG Fellows Program at Harvard Medical School. Machine Learning for Discovering Phenotypes in Autism Spectrum Disorders.
- Speaker, May 2013, Northeastern University. Bayesian Nonparametric Methods for Timeseries Analysis.
- Speaker, March 2013. Vecna. Characterizing Temporal Patterns in Autism Spectrum Disorder from Electronic Health Records.

- Speaker, September 2012. Institute for Computational and Experimental Research in Mathematics at Brown University. Bayesian nonparametric Approaches to Reinforcement Learning in Partially Observable Domains.
- Speaker, November 2011. Tufts University, Computer Science Seminar Series. Bayesian Nonparametric Approaches to Reinforcement Learning in Partially Observable Domains.
- Guest Lecturer, October 2011, Harvard University. CS281: Advanced Machine Learning (advanced graduate course). Prepared and presented a pair of lectures on inference in time series.
- Speaker, July 2011. Wellesley College, Summer Seminar Series. Towards Better AI: Using Bayesian Nonparametric Methods to Build More Flexible Agents.
- Guest Lecturer, June 2011, Planning Under Uncertainty Course; Lincoln Labs. Bayesian Reinforcement Learning for Dialog Management.
- Speaker, April 2011. University of Toronto, Machine Learning Seminar. Bayesian Nonparametric Approaches to Reinforcement Learning in Partially Observable Domains.
- Speaker, March 2011. Rutgers University, Rutgers Laboratory for Real-Life Reinforcement Learning Group. Bayesian Nonparametric Approaches to Reinforcement Learning in Partially Observable Domains: The Nuts and Bolts.
- Guest Lecturer, Fall 2010, Massachusetts Institute of Technology. 16.420: Planning Under Uncertainty (advanced graduate course). Prepared and presented a pair of lectures on time series analysis and Bayesian model learning.
- Speaker, November 2010, University of Massachusetts Amherst: Bayesian Nonparametric Approaches to Reinforcement Learning in Partially Observable Domains.
- Speaker, April 2010, Brown University. Efficient Inference for the Indian Buffet Process.

Honors and Publicity

- Faculty:
  - Anita Borg Early Career Award 2021
  - Commendation for Extraordinary Teaching for CS181 2020
  - Harvard Graduate Student Council Everett Mendelsohn Excellence in Mentoring Award 2019
  - Harvard SEAS Capers and Marion McDonald Mentoring Award 2019
  - Sloan Fellow 2018
  - NSF CAREER Award 2018
  - Google Faculty Research Award 2018
  - Air Force Young Investigator Program Award 2016
  - Early Career Spotlight IJCAI 2016
  - Stanford-Whitehouse Symposium Invited Speaker 2016
  - Publicity for Explanation and the Law
    - Contributed to the OECD report “Artificial Intelligence in Society” (June 2019)
    - Contributed to Academy of Medical Royal Colleges report “Artificial Intelligence in Healthcare” (January 2019)
    - TEDx Talk featured on Talking Machines by Katherine Gorman (December 2018)
    - OpEd in the Washington Post by FDV and (March 2018)
    - News in the Communications of the ACM by Don Monroe (November 2018)
    - Harvard Magazine by Jon Shaw (December 2018)
    - Parts incorporated into recommendations to the German Digital Council (reports to Angela Merkel) by Urs Gasser (November 2018)
- Postdoctoral: Named one of IEEE's AI “10 to Watch” (2013), NSF CiTraCS Postdoctoral Fellow (2012), Invited participant at the Rising Stars in EECS workshop, organized by MIT (2012)
- Graduate: Hugh Hampton Young Fellowship recipient (2010-2011), Marshall Scholar (2006), Trinity Prince of Wales Research Student, Trinity College External Research Studentship (2007), NDSEG Graduate Fellow (2005), NSF Graduate Fellow (2005)
- Undergraduate: Honor Society Memberships: Phi Beta Kappa, Sigma Gamma Tau (aerospace honors society), Sigma Pi Sigma (physics honor society); Goddard Research Award for best research project at the NASA Academy (2004), Manufacturing Award for undergraduate junior-senior design project (2005), Athletics Gold Award for Service for founding Club Sports Council (2005), Todd Anderson Teaching Award for excellence in teaching in the MIT Experimental Studies Group (2005).
- High School: Many, including FIRST robotics competition and the state chemistry olympiad, but the one I am most proud of is 1st place at Destination Imagination Global Finals, 2001.

Early Teaching Experience

- **Instructor** Winter 2010-2012, Summer 2011
  Massachusetts Institute of Technology. 6.085: Statistics for Research Projects. Developed and taught a half-course on statistics for undergraduate researchers. Teaching rated 6.3/7.0 (departmental average is 5.3).
- **Program Director** Winter 2005, 2006
  Massachusetts Institute of Technology. 6.185: Mobile Autonomous Systems Laboratory. Directed a completely volunteer, student-run autonomous vision-based navigation robotics contest. Duties included organizing the production of student kits, raising $25K, designing curriculum, and managing a 10-person teaching staff in addition to teaching and lab duties (also staff in 2007).
- **Instructor** Fall 2005, 2006
  Massachusetts Institute of Technology. 18.01: Single Variable Calculus; 18.02: Multivariable Calculus.
- **Supervisor** (teaching assistant) Fall 2007, 2008; Winter 2008
  University of Cambridge. Tutoring, practicals for undergraduate signals and systems and statistical modeling.
- **Teaching Assistant** Spring 2002 – Fall 2006, Spring 2007
  Massachusetts Institute of Technology. Undergraduate Courses: 18.01: Single Variable Calculus, 18.02: Multivariable Calculus, 18.03: Differential Equations, 8.022 Electricity and Magnetism; Graduate Course: 6.437 Inference and Information. Prepared tutorials, office hours. Teaching for 6.437 rated 6.3/7.0 (departmental average is 5.6): "was highly praised as an amazing TA. Students complemented her knowledge of the field, great blackboard techniques, and clear explanations."

Early Projects Supervised

- Project Mentor, Winter 2013—Fall 2014. Massachusetts Institute of Technology. Supervised two undergraduate students to apply basic machine learning techniques to predict diagnoses in autism spectrum disorders and inflammatory bowel disease.
- Bachelors Thesis Mentor, Winter 2012—Fall 2014, Wellesley College. Co-supervised two undergraduate theses to predict energies of solvation based on molecular features.
- Project Mentor, Fall 2012-Spring 2013. Harvard. Supervised a project to classify pot-holes based on smart phone accelerometer data.
- Project Mentor, Summer 2012, Harvard Medical School. Supervised an undergraduate project looking at pathways with high differential gene expression between autistic individuals and a healthy population.
- Project Mentor, Summer 2010, Massachusetts Institute of Technology. Helped supervise an undergraduate project for localizing wheelchairs using wifi signal strengths.
- Masters Thesis Mentor, Fall 2008 – Spring 2009, University of Cambridge. Helped supervise a masters thesis on learning Bayesian network structure with large missing data and an application to water and sanitation.
- Project Mentor, Spring 2007, Massachusetts Institute of Technology. Helped supervise a junior-senior design project in the effectiveness of wheelchair dialog managers.

Early University and Departmental Service

- Committee on Foreign Scholarships, Massachusetts Institute of Technology, 2009-Present. Coached students applying for foreign scholarships, participated in mock interviews.
- REF, EECS Department, MIT Summer 2010-Summer 2012. Peer mediator, conflict coach, and stress reduction resource for graduate students. Certification includes 32 hours of mediation training.
- Graduate leadership: Massachusetts Institute of Technology: Graduate Resident Tutor 2006-2007. EECS Orientation Co-chair 2006. Grad Women in EECS Discussion Coordinator 2007. Program chair of CSAIL Student Workshop 2007. EECS Academic Chair 2010. Organizer/Founder of Robotics Tea 2010. University of Cambridge: Co-founder of Engineering Students Executive Committee 2007-2008. Group Social Chair 2007-2009.
- Undergraduate leadership: Massachusetts Institute of Technology: Officer, Club Sports Council, 2002-2003 (Athletics Department Gold Award for Service); Taekwondo Club President 2002-2005. Hall Chair 2004-2005.

Early Community Outreach

Statistical Consulting

- Volunteer Analyst, Statistics without Borders. 2009—present. Projects: Re-analyzed data from a study on the effectiveness of maternal health education on social and health outcomes for CARE in Bangalore, India (Winter 2011). Analyzed influences on choices surrounding the consumption of animal products for the Farm Sanctuary (Fall 2013).
- Volunteer Analyst, Learning Unlimited, 2011-2012. Organized focus groups to determine what aspects of MIT's Educational Studies Program's events had the most lasting effects on students.
- Expert reviewer for 'A Preliminary Assessment of the Impact of the Liberian Truth and Reconciliation Commission,' an independent evaluation undertaken as part of the Human Rights Data Analysis Group (HRDAG) at Benetech, January 2010.
- Statistical consultant for a Millennium Villages project on evaluating effectiveness of various diarrhea treatment programs in Africa, Winter 2010.

Educational Outreach

- Speaker, Maggie L. Walker Governor's School, Spring 2011. Invited talk for the Mu Alpha Theta Math Honors Society on Bayesian Nonparametric Methods for Reinforcement Learning; guest lecturer for the Linear Algebra and Math Modeling classes.
- Speaker, MIT Women's Initiative, Winter 2011, Massachusetts Institute of Technology.
- Speaker, Young Jain Convention Summer 2010, 2012.
- Educational Studies Program, Fall 2002—2012. Massachusetts Institute of Technology. Co-chair, Treasurer, Publicity Chair. Directed SAT prep and enrichment programs for 1000+ students and managed $60K budgets. Taught classes for up to 250 students ranging from statistics to story-telling.
- Intern, Science Museum of Virginia, Richmond VA, Summer 2003.

Early Research and Industry Experience

- **Bioinformatics Analyst** April 2012 – August 2012
  Brigham and Women's Hospital, Boston MA. Applied machine learning techniques to characterize cardiovascular risk among healthy women.
- **Research Assistant** September 2005 – December 2012
  Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Developed techniques for learning parameters of dialog managers online (masters thesis). Designed Bayesian nonparametric models and inference techniques for reinforcement learning applications (doctoral thesis).
- **Research Assistant** October 2007 – August 2009
  Computational and Biological Learning Laboratory, University of Cambridge. Developed efficient inference techniques for the Indian Buffet process.
- **Intern** Summer 2007
  ITA Software. Boston, MA. Applied machine learning techniques for econometric analysis of flight data.
- **Undergraduate Research Projects** February 2002 – June 2005
  Massachusetts Institute of Technology. Completed a variety of projects, including designing an alternative keyboard (self-directed); testing fault-detection software for autonomous rovers (Model-based Embedded Robotic Systems Group); designing fly-wheel circuitry (Space Systems Lab); creating a gesture recognition interface (self-directed); and improving video alignments for robot localization (Vision Group).
- **Research Associate** Summer 2004
  NASA Academy, NASA Goddard. Developed algorithms for minimum-fuel spacecraft rendezvous.
- **Lab Assistant** Summer 2003
  Sentor Technologies Inc., Richmond, VA. Programmed and built control circuitry for a gamma ray detector.