jzf2101@columbia.edu
jzf2101

Jessica Zosa Forde

Education

PhD, *Brown University*, Computer Science, Expected 2024.

MA, *Columbia University*, Statistics, 2014.

BA, *Wellesley College*, Political Science, 2008.

Professional Experience

Facebook AI Research

Research Intern. 2019, 2020–Present.

Project Jupyter

Core Maintainer, JupyterHub and JupyterLab Team. 2017–2019.

Tufts University

Lecturer, Gordon Institute. 2019.

Lawrence Berkeley National Laboratory

Affiliate. 2018–2019.

University of California, Berkeley

Berkeley Institute of Data Science Visiting Researcher. 2017–2019.

New York University

Visiting Researcher, Kyunghyun Cho Group. 2017.

Harvard University

Visiting Fellow, Finale Doshi-Velez Group. 2017.

Careful

Data Science Consultant. 2016–2017.

Qadium

Data Scientist, DARPA XDATA Funded Datamicroscopes Project. 2015.

PRO Unlimited for McKinsey & Company

Data Scientist Consultant, People Insights. 2015.

Columbia University

Machine Learning Consultant, Center for Computational Learning Systems (CCLS) Di-BOSS Project. 2013–2014.

Publications

Submitted for Publication

M. Paganini and **J. Z. Forde**, "dagger: A Python Framework for Reproducible Machine Learning Experiment Orchestration,"

M. Paganini and **J. Z. Forde**, "Bespoke vs. Prêt-à-Porter Lottery Tickets: Exploiting Mask Similarity for Trainable Sub-Network Finding,"

J. Zech*, **J. Z. Forde***, and M. Littman, "Individual predictions matter: Assessing the effect of data ordering in training fine-tuned CNNs for medical imaging,"

Journal Papers

J. Zech, **J. Forde**, J. Titano, D. Kaji, A. Costa, and E. K. Oermann, "Detecting insertion, substitution, and deletion errors in radiology reports using neural sequence-to-sequence models," *Annals of Translational Medicine*, 2018.

Conference Papers

Project Jupyter, M. Bussonnier, **J. Forde**, J. Freeman, B. Granger, T. Head, C. Holdgraf, K. Kelley, G. Nalvarte, A. Osheroff, M. Pacer, Y. Panda, F. Perez, B. Ragan-Kelley, and C. Willing, "Binder 2.0 - Reproducible, interactive, sharable environments for science at scale," in *Proceedings of the 17th Python in Science Conference (SciPy)*, Austin, Texas, 2018, pp. 113–120.

Workshop Papers

M. Paganini and **J. Forde**, "Streamlining Tensor and Network Pruning in PyTorch," in *Practical ML for Developing Countries Workshop at ICLR*, **Contributed Talk**, 2020.

M. Paganini and **J. Z. Forde**, "On iterative neural network pruning, reinitialization, and the similarity of masks," in *Practical Machine Learning for Developing Countries ICLR Workshop*, 2020.

J. Z. Forde and M. Paganini, "The Scientific Method in the Science of Machine Learning," in *Debugging Machine Learning Models ICLR Workshop*, **Contributed Talk**, 2019.

J. Z. Forde, J. Zech, and M. Littman, "Assessing variability in predictions of a deep learning model to identify findings in chest radiographs," in *Machine Learning For Health Workshop, NeurIPS*, 2019.

J. Z. Forde, J. Zech, and M. Littman, "Individual predictions matter: An example from deep learning for medical imaging," in *Science Meets Engineering NeurIPS Workshop*, 2019.

J. Forde, "Code as Scholarship: Extensible Software Experiments," in *Critiquing and Correcting Trends in Machine Learning NeurIPS Workshop*, **Spotlight Talk**, 2018.

J. Forde, M. Bussonnier, F.-A. Fortin, B. Granger, T. Head, C. Holdgraf, P. Ivanov, K. Kelley, M. Pacer, Y. Panda, F. Perez, G. Nalvarte, B. Ragan-Kelley, Z. Sailer, S. Silvester, E. Sundell, and C. Willing, "Reproducing Machine Learning Research on Binder," in *Machine Learning Open Source Software 2018: Sustainable communities NeurIPS workshop*, 2018.

J. Forde, T. Head, C. Holdgraf, Y. Panda, G. Nalvarte, B. Ragan-Kelley, and E. Sundell, "Reproducible Research Environments with Repo2Docker," in *Reproducibility in ML Workshop, ICML*, 2018.

A. S. Ross and **J. Forde**, "Refactoring Machine Learning," in *Critiquing and Correcting Trends in Machine Learning NeurIPS Workshop*, 2018.

J. Forde and L. A. Hannah, "A reinforcement learning solution for steam heating decision making," in *Women in Machine Learning Workshop*, 2013.

Conference Abstracts

J. Zech, **J. Forde**, J. Titano, D. Kaji, A. Costa, and E. K. Oermann, "Detecting insertion, substitution, and deletion errors in radiology reports using neural sequence-to-sequence models," in *Machine Learning for Healthcare Conference*, 2018.

Patent Applications

R. N. Anderson, A. Boulanger, V. Bhandari, E. Boniberger, A. Gagneja, J. Gilbert, A. Kressner, A. Rajan, D. Solomon, **J. Forde**, L. L. Wu, V. Rathod, K. Morenski, and H. Shokri, "Total property optimization system for energy efficiency and smart buildings," 20150178865:A1, 2015.

Technical Reports

J. Forde, C. Holdgraf, Y. Panda, M. B. Aaron Culich, M. Ragan-Kelley, M. Pacer, C. Willing, T. Head, F. Perez, B. Granger, and Project Jupyter Contributors, "Post-training evaluation with Binder," in *Conference on Fairness, Accountability, and Transparency.*, 2018.

Project Jupyter Contributors, C. Titus Brown, M. Bussonnier, **J. Forde**, J. Freeman, B. Granger, T. Head, C. Holdgraf, M. Pacer, Y. Panda, F. Perez, M. Ragan-Kelley, and C. Willing, "Introducing Binder 2.0 — share your interactive research environment," in *eLife Labs*, eLife Sciences Publications Limited, 2017.

Demonstrations

J. Z. Forde, M. Bussonnier, F.-A. Fortin, B. Granger, T. Head, C. Holdgraf, P. Ivanov, K. Kelley, F. Perez, M. Pacer, Y. Panda, G. Nalvarte, M. Ragan-Kelley, Z. Sailer, S. Silvester, E. Sundell, and C. Willing, "Reproducing Machine Learning Research on Binder," in *NeurIPS Demo Track*, 2018.

J. Forde, V. Rathod, H. Shookri, V. Bandari, A. Rajan, J. Min, A. Fan, L. Wu, A. Gagneja, D. Riecken, D. Solomon, L. Hannah, A. Boulanger, and R. Anderson, "Di-BOSS™: Digital building operating system solution," in *NeurIPS Demo Track*, 2013.

Open Source Maintainer

Project Jupyter

JupyterLab: an extensible environment for interactive computing. 2017–present.

JupyterHub: a multi-user Hub which spawns, manages, and proxies multiple instances of the single-user Jupyter notebook server. 2017–present.

BinderHub: `BUILD` and `REGISTER` a Docker image using a GitHub repository, then `CONNECT` with a Kubernetes JupyterHub. 2017–present.

repo2docker: fetches a git repository and builds a Docker container image based on the configuration files found in the repository. 2017–present.

IPython: interactive computing in Python. 2018–present.

Jupyter Docker Stacks: ready-to-run Docker images containing Jupyter applications and interactive computing tools. 2018–present.

nteract project

nteract: stand-alone desktop application for computational notebooks. 2018-present.

Qadium

datamicroscopes: Bayesian nonparametric models in Python. 2015.

Columbia University

Density: university population estimates based on wireless connection data. 2014-2016.

Awards

NumFOCUS Contributor Award. 2018.

Press

"Project Jupyter with Jessica Forde, Yuvi Panda and Chris Holdgraf." Google Cloud Platform Podcast, Episode 122, 2018.

Presentations

Invited Talks

"Promoting Science in Machine Learning Research with Binder." *Facebook AI Research.* Menlo Park. 2019.

"Promoting Science in Machine Learning Research." *Deep Learning for Science School.* Lawrence Berkeley National Laboratory. 2019.

"Promoting Science in Machine Learning Research." *Reproducibility in Machine Learning ICLR workshop.* New Orleans. 2019.

"Doubling Down on the Scientific Method." *Data, Learning and Inference (DALI) Healthcare in Machine Learning Workshop.* George, South Africa. 2019.

"Code as Scholarship: Extensible Software Experiments." Humanity-Centered Robotics Initiative, Brown University. 2018.

"Extensible Machine Learning Research." Reasoning and Learning Lab, McGill University. 2018.

"On Extensibility." Data to Actionable Knowledge Group, Harvard University. 2018.

"Tools for Reproducibility and Extensibility in Scientific Research." Laboratory for Computational Physiology, MIT. 2018.

"JupyterLab, Building Blocks for Interactive Computing." National Energy Research Scientific Computing Center (NERSC) Data Seminar, Lawrence Berkeley National Laboratory. 2018.

"Tools for Reproducibility and Extensibility in Scientific Research." Council for Scientific and Industrial Research (CSIR). Pretoria, South Africa. 2018.

"Tools for Reproducibility and Extensibility in Scientific Research." CERN EP/IT Data Science Seminar. Geneva, Switzerland. 2018.

"JupyterLab: Building Blocks for Interactive Computing." New York Python Meetup, Two Sigma. 2018.

"On Jupyter." Thoughtbot. New York, NY. 2018.

"An Introduction to Open Source." Application Development Intiative, Columbia University. 2017.

"An Introduction to Reinforcement Learning." New York Python Meetup, Flatiron Health. 2017.

"An Urban Introduction to Reinforcement Learning." Thoughtbot. New York, NY. 2015.

Invited Panels

Data, Learning and Inference Healthcare (DALI) in Machine Learning Workshop. George, South Africa. 2019.

Tutorials

"Zero To JupyterHub with Kubernetes." *PyBay*, Bloomberg. 2018.

"Build a Binder." with Tim Head *University of Birmingham.* Birmingham, UK. 2018.

"Hands-on Tutorial Deploying Portable Research and Learning Environments on Commercial Clouds and XSEDE." with Aaron Culich, Félix-Antoine Fortin, Chris Hench, Chris Holdgraf, M Pacer, Carol Willing, Yuvi Panda, and Min Ragan-Kelley. *Practice & Experience in Advanced Research Computing (PEARC).* 2018.

"JupyterLab, Building Blocks for Interactive Computing." CERN EP/IT Data Science Seminar. Geneva, Switzerland. 2018.

Conference Presentations

"Code as Scholarship: Extensible Software Experiments." *Critiquing and Correcting Trends in Machine Learning NeurIPS Workshop.* Spotlight Talk. 2018.

"Tools for Reproducibility and Extensibility in Scientific Research." *Future of Research Communications and e-Scholarship (FORCE11).* McGill University. 2018.

"Learn by doing: Using data-driven stories and visualizations in the (high school and college) classroom." with Carol Willing and Erik Sundell. *JupyterCon*, New York, NY. 2018.

"Detecting insertion, substitution, and deletion errors in radiology reports using neural sequence-to-sequence models." *Machine Learning for Healthcare Conference.* Spotlight Talk. 2018.

"JupyterLab, Building Blocks for Interactive Computing." *PyCaribbean*, Santo Domingo, Dominican Republic. 2018.

"Party Planning in Python (An Introduction to Optimization Solvers.)" *PyData New York*, Microsoft. 2017.

"JupyterLab, Building Blocks for Interactive Computing." with Jason Grout *Strata Data Conference*, New York, NY. 2017.

"An Introduction to Reinforcement Learning." *PyCon*, Portland Oregon. 2017.

"An Introduction to Reinforcement Learning." *PyGotham*, United Nations. 2016.

"A Brief Introduction to Reinforcement Learning." *AI Camp*, United Nations. 2016.

"Datamicroscopes: Bayesian nonparametric models in Python." *DataPoint NYC*, Google, 2015.

"Visualizing Wireless-Router Timeseries Data with the Density API, Seaborn, and Pandas." *PyData NYC*, Bank of America. 2015.

Poster Presentations

"JupyterHub Tools for Reproducibility and Extensibility." *Deep Learning Indaba*, Stellenbosch University, South Africa, 2018.

"Post-Training evaluation with repo2docker." *Self Organizing Machine Learning Conference*, Google. 2017.

"A Reinforcement Learning Solution for Steam Heating Decision Making." *Women in Machine Learning Workshop*. 2013.

Service

Organizing Committee

I can't believe it's not better! Bridging the gap between theory and empiricism in probabilistic machine learning workshop, NeurIPS. 2020.

Retrospectives NeurIPS Workshop. 2020.

Retrospectives ICML Workshop. 2020.

NeurIPS Reproducibility Challenge. 2019.

Retrospectives NeurIPS Workshop. 2019.

Area Chair

Women in Machine Learning Workshop. 2018, 2019.

Program Committee

Machine Learning and Systems Reproducibility Initiative. 2020.

Machine Learning for Health NeurIPS Workshop. 2018, 2019.

Machine Learning the Physical Sciences NeurIPS Workshop. 2019.

Latinx in AI. 2019.

Grant Reviewer

Ethics and Governance of AI Initiative Challenge on AI and the News. 2018.

Reviewer

NeurIPS. 2015, 2020.

ACM CHIL. 2020.