

EDUCATION

- Harvard University** (Boston, MA) *MSc, PhD in Biostatistics* May 2019
Research Keyword: Bayesian Machine Learning, Ensemble Learning, Uncertainty Quantification, Robust Statistics
- University of Iowa** (Iowa City, IA) *BS in 1Statistics, Mathematics, Minor Computer Science* May 2013
magna cum laude

PROFESSIONAL EXPERIENCE

Google Research

Staff Research Engineer 2022-Pres

- Driving research efforts and the adoption of principled methods for large-scale decision making systems under noise, bias, and uncertainty, with broad applications in language understanding and predictive modeling.

Senior Research Engineer 2019-2022

- Developed statistical solutions to fundamental issues in artificial intelligence (uncertainty quantification and decision making), with application to language processing and recommender systems.
- Primary appointment under Language division of Google Research, in close collaboration with Google Brain.

Research Intern 2018

- Project focused on genomic mutation (i.e. structural variant) detection using deep learning methods . Work under Google Accelerated Science, in close collaboration with Google Brain Genomics.
- Developed a novel neural network module to perform specialized, vision-based processing of gene-sequencing information. Illustrated significant accuracy improvement on mutation type detection tasks.
- Spearheaded the design and implementation of a deep-learning-based system (main architecture: multitask resnet with self-attention) to perform streamlined feature-extraction, mutation site detection and mutation type classification. Illustrated precision and recall improvement over existing structural variant detection tools.

Department of Biostatistics, Harvard University 2019-2021

Visiting Scientist

- Develop rigorous statistical/machine learning methodology for (1) uncertainty quantification for air pollution exposure assessment, and (2) health effect estimation in large-scale environmental health studies.

Martinos Center for Biomedical Imaging, Mass General Hospital

Department of Data Science, Dana Farber Cancer Institute 2017-2019

Graduate Research Fellow / Machine Learning Scientist

- Conducting computational study for evaluating Bayesian inference algorithms for microbiome factor modeling.
- Building reinforcement learning system for automated discovery of novel MRI configurations.
- Participated in theory development and design of manifold-inspired deep learning architecture for MRI image reconstruction (*Nature* vol 555).

learnable.ai 2017-2018

Lead Research Engineer

- Designed and supervised the implementation (leading four software engineers) of the company’s optical character recognition (OCR) pipeline for processing whole-page mathematical documents.
- Developing a system (leading two research engineers) for joint vision- and language-based understanding and reasoning for high-school geometry questions.
- Provided technical guidance and helped design R& D agenda for classroom video/audio understanding pipeline.
- Other duties include reviewing relevant literature and plan technical solutions, designing and executing R& D agenda, supervising engineer/research progress, and mentoring/management of machine learning engineer interns.

Harvard Clean Air Research Center
Assistant Statistician

2013-2015

- Built spatiotemporal prediction system for heavy-metal air pollutants by integrating information from various sources (air monitoring records, meteorological information, etc) under Random Forrest and Kernel Regression.
- Implemented automated feature selection for GIS features using a combination of measurement error-based weighting and Ridge-type penalization. Conducted stratified cross validation to assess the model’s out-of-sample prediction and the influence of prediction error on the risk estimation in second-stage association studies.

TECHNICAL SKILLS

- **Analysis & Modelling:** Python (tensorflow, pytorch, pyMC3), R, Matlab
- **Graphics & Documents:** ggplot2, OpenGL, Shiny, ArcGIS, L^AT_EX
- **High Performance Computing:** C (CUDA, OpenCL, OpenMP)
- **Software Development:** Python, C++, Java, Bash

MENTORSHIP EXPERIENCE

Zi Lin, Doctoral Candidate in Computer Science, UCSD

2020-Pres.

- **Project 1:** Pruning Redundant Mappings in Transformer Models via Spectral-Normalized Identity Prior. *Findings of ACL 2021*
- **Project 2:** Towards Collaborative Neural-Symbolic Graph Semantic Parsing via Uncertainty. *Findings of ACL 2022*
- **Project 3:** Neural-Symbolic Inference for Robust Autoregressive Graph Parsing via Compositional Uncertainty Quantification. *EMNLP 2022 (To Appear)*.

Wenyng Deng, Doctoral Candidate in Biostatistics, Harvard University

2018-Pres.

- **Project 1:** A Bootstrap Test for Nonlinear Interaction using Cross-validated Kernel Ensemble. *arXiv:1811.11025*
- **Project 2:** Towards a Unified Framework for Uncertainty-aware Nonlinear Variable Selection with Theoretical Guarantees. *NeurIPS 2022 (To Appear)*

Beau Coker, Doctoral Candidate in Biostatistics, Harvard University

2019-Pres.

- **Project 1:** High-dimensional Neural Mediation Analysis for Nonlinear Continuous Treatments. *In Progress*

PROFESSIONAL SERVICE

Review Committee, NeurIPS 2020-2022, ICML 2020-2022, ICLR 2019-2023. AISTATS 2022-2023. Machine Learning, JMLR, TMLR, IJUQ.

Program Committee, ICML 2020 Workshop on Uncertainty & Robustness in Deep Learning, NeurIPS 2021 Workshop on Distribution Shifts.

SELECTED PUBLICATIONS

Machine Learning, Theory & Method

- JZ Liu**, Padhy, S., Ren, J., Lin, Z., Wen, Y., Jerfel, G., Nado, Z., Snoek, J., Tran, D. and Lakshminarayanan, B. *A Simple Approach to Improve Single-Model Deep Uncertainty via Distance-Awareness*. Journal of Machine Learning Research. (To Appear)
- W Deng, B Coker, R Mukherjee, **JZ Liu*** (Co-senior Author), B Coull*. *Towards a Unified Framework for Uncertainty-aware Nonlinear Variable Selection with Theoretical Guarantees*. Proceedings of 36th Conference on Neural Information Processing Systems (NeurIPS 2022) (To Appear) (Spotlight Presentation, ICML 2022 Workshop in Interpretable ML in Healthcare (IMLH))
- Liu JZ**. *Variable selection with rigorous uncertainty quantification using deep bayesian neural networks: Posterior concentration and bernstein-von mises phenomenon*. Proceedings of The 24th International Conference on Artificial Intelligence and Statistics (AISTAT 2021)
- M Havasi, R Jenatton, S Fort, **JZ Liu**, J Snoek, et al. *Training independent subnetworks for robust prediction*. Proceedings of 2020 International Conference on Learning Representations (ICLR 2020).
- Liu JZ**, Lin Z, Padhy S, Tran D, Bedrax Weiss T, Lakshminarayanan B. *Simple and principled uncertainty estimation with deterministic deep learning via distance awareness*. Advances in Neural Information Processing Systems 33 (NeurIPS 2020)
- Liu JZ**, Paisley J, Kioumourtzoglou M, Coull B. *Accurate Uncertainty Estimation and Decomposition in Ensemble Learning*. Advances in Neural Information Processing Systems 32 (NeurIPS 2019)
- Liu JZ**, Paisley J, Kioumourtzoglou M, Coull B. *Adaptive and Calibrated Ensemble Learning with Tail-free Process*. Bayesian Nonparametrics workshop, NeurIPS 2018.
- Liu JZ**, Coull B. *Robust Hypothesis Test for Nonlinear Effect with Gaussian Processes*. Advances in Neural Information Processing Systems 30 (NeurIPS 2017)

Machine Learning, Application

- Z Lin, **J Liu*** (Co-senior Author), J Shang*. *Neural-Symbolic Inference for Robust Autoregressive Graph Parsing via Compositional Uncertainty Quantification*. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022) (To Appear)
- ID Kivlichan*, Z Lin*, **J Liu*** (Co-first Author), L Vasserman. *Measuring and Improving Model-Moderator Collaboration using Uncertainty Estimation*. Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)
- Liu JZ**, Lee J, Lin P, Valeri L, Christiani D, Bellinger D, Wright R, Mazumdar M, Coull B *A Robust Hypothesis Test for Continuous Nonlinear Interactions in Nutrition-Environment Studies: A Cross-validated Ensemble Approach*. Journal of the American Statistical Association. (Distinguished Paper Award, ENAR 2019)
- Zhu B, **Liu JZ**, Rosen B, Rosen M *Image reconstruction by domain transform manifold learning*. Nature Vol 555, (22 March 2018) doi:10.1038/nature25988
- Zhu B, **Liu J**, Koonjoo N, Rosen B, and Rosen M *AUTOMated pulse SEquence generation (AUTOSEQ) using Bayesian reinforcement learning in an MRI physics simulation environment*. Joint Annual Meeting ISMRM-ESMRMB 2018
- Deng W, **Liu JZ**, E Lake, B Coull. *CVEK: Robust Nonlinear Effect Estimation and Testing with Gaussian Process Ensemble*. Journal of Statistical Software. *arXiv:1811.11025*

Public Health & Biomedicine

- H Li, W Deng, R Small, J Schwartz, **J Liu*** (Co-senior Author), L Shi* *Health effects of air pollutant mixtures on overall mortality among the elderly population using Bayesian kernel machine regression (BKMR)*. Chemosphere 286 (2022), 131566.
- Hswen Y, Brownstein J, **Liu JZ**, Hawkins J *Use of a Digital Health Application for Influenza Surveillance in China*. American Journal of Public Health, 2017; e1 DOI: 10.2105/AJPH.2017.303767

Wang Z, Zheng Y, Zhao B, Zhang Y, **Liu Z**, Xu J, Chen Y, Yang Z, Wang F, Wang H, He J, Zhang R, Abliz Z. *Human Metabolic Responses to Chronic Environmental Polycyclic Aromatic Hydrocarbon Exposure by a Metabolomic Approach*. Journal of Proteome Research, 2015, 14 (6), pp 2583 - 2593

Liu Z, Zhang J, Zhao B, et al. *Population-based reference for birth weight for gestational age in northern China*. Early Human Development 2014;90(4):177-87.

HONORS & AWARDS

IMS Hannan Travel Award, Institute of the Mathematical Statistics, 2019

ENAR Distinguished Paper Award, International Biometric Society, 2019

Certificates of Distinction and Excellence in Teaching, Harvard Derek Bok Center for Teaching and Learning, 2018

Phi Beta Kappa, Alpha of Iowa Chapter, CLAS, University of Iowa, 2012