Mikhail Yurochkin

moonfolk@umich.edu • moonfolk.github.io

Education

Ph.D., Statistics 2013 – 2018
University of Michigan, Ann Arbor, MI, United States
- Thesis: "Geometric Inference in Bayesian Hierarchical Models with Applications to Topic Modeling"
- Advisor: Professor XuanLong Nguyen

Master of Arts, Statistics 2013 – 2015
University of Michigan, Ann Arbor, MI, United States

Bachelor's, Applied Mathematics and Physics 2009 – 2013
Moscow Institute of Physics and Technology, Moscow, Russia

Research Experience

Research Manager Nov 2023 – Present
Research Staff Member June 2018 – Present
IBM Research and MIT-IBM Watson AI Lab, Cambridge, MA, United States
Statistical Large Language Modeling group, Team Lead
- Leading research projects supporting IBM's platform for Large Language Models.
- Leading research teams on Algorithmic Fairness, Federated Learning, Out-of-Distribution Generalization, and Optimal Transport projects. Our work is regularly published in top-tier Machine Learning and Artificial Intelligence venues.
- Leading a team of engineers to create AI Fairness software and demonstrations.
- Mentoring graduate students and collaborating with academic research groups, including Yuekai Sun's group at the University of Michigan, Justin Solomon's Geometric Data Processing group at MIT, and Anette (Peko) Hosoi's group at MIT.
- Regularly representing IBM Research at conferences and workshops via invited talks and tutorial presentations.

Research Associate May 2018 – June 2018
University of Michigan, Ann Arbor, MI, United States
- Led two research projects in collaboration with senior PhD students and faculty resulting in two NeurIPS 2018 submissions.
- Studied driving behaviors using Bayesian modeling in collaboration with the Toyota Research Institute.

Data science research intern June 2017 – Aug 2017
Adobe, San Jose, CA, United States
- Proposed a method for jointly training a graph neural network and inferring the graph structure when the graph is unknown. Applied this method to forecasting number of visits to a major retailer's website across cities for improved marketing and sales.

Consultant for Statistics, Computing, and Analytics Research Sept 2016 – Dec 2017
University of Michigan, Ann Arbor, MI, United States
- Assisted faculty and graduate students in areas such as biology, dentistry, marketing, political science, public health, and computer science in identifying and implementing appropriate statistical methodology for their data and research problems.

Science team intern May 2016 – Aug 2016
LogicBlox/Predictix, Atlanta, GA, United States
- Proposed and implemented an explainable forecasting model based on Factorization Machines and Indian Buffet Process. Analyzed sales data from a major retailer using the proposed model and published a NeurIPS 2017 paper based on this work.

Graduate Student Research Assistant July 2014 – April 2018
University of Michigan, Ann Arbor, MI, United States
- Formulated topic modeling as a geometric problem and developed new algorithms significantly outperforming prior state-of-the-art in terms of estimation speed. These results were published at NeurIPS 2016 and 2017.
- Proposed Bayesian modeling approaches to identify insider attacks from SQL queries of a bank database.

Teaching Experience

Guest Lecturer on Individual Fairness @ UT El Paso, Responsible AI Class — March 2024
Guest Lecturer on Fairness in ML @ University of Michigan, Introduction to ML Class — Nov 2022
Graduate Student Instructor — Sept 2013 – April 2016
University of Michigan, Ann Arbor, MI, United States
Taught labs, graded, held office hours:

- Topics in Biostatistics — Jan 2015 – April 2016
- Applied Probability — Sept 2014 – Dec 2014
- Introduction to Statistics — Jan 2014 – April 2014
- Introduction to Probability — Sept 2013 – Dec 2013

Publications

[1] F. M. Polo, L. Weber, L. Choshen, Y. Sun, G. Xu, and **M. Yurochkin**. tinyBenchmarks: evaluating LLMs with fewer examples. *ICLR Workshop on Mathematical and Empirical Understanding of Foundation Models (ME-FoMo)*, 2024.

[2] L. Ngweta, M. Agarwal, S. Maity, A. Gittens, Y. Sun, and **M. Yurochkin**. Aligners: Decoupling LLMs and Alignment. In *Tiny Papers Track at the International Conference on Learning Representations (ICLR)*, 2024. **Notable**.

[3] H. Wang, F. Polo, Y. Sun, S. Kundu, E. Xing, and **M. Yurochkin**. Fusing Models with Complementary Expertise. In *International Conference on Learning Representations (ICLR)*, 2024.

[4] S. Maity, M. Agarwal, **M. Yurochkin**, and Y. Sun. An Investigation of Representation and Allocation Harms in Contrastive Learning. In *International Conference on Learning Representations (ICLR)*, 2024.

[5] F. Petersen, A. Mishra, H. Kuehne, C. Borgelt, O. Deussen, and **M. Yurochkin**. Uncertainty Quantification via Stable Distribution Propagation. In *International Conference on Learning Representations (ICLR)*, 2024.

[6] T. Shnitzer, A. Ou, M. Silva, K. Soule, Y. Sun, J. Solomon, N. Thompson, and **M. Yurochkin**. Large Language Model Routing with Benchmark Datasets. *NeurIPS Workshop on Distribution Shifts (DistShift)*, 2023. **Oral**.

[7] Y. Zeng, K. Greenewald, L. Jung, K. Lee, J. Solomon, and **M. Yurochkin**. Outlier-Robust Group Inference via Gradient Space Clustering. *NeurIPS Workshop on Distribution Shifts (DistShift)*, 2023.

[8] R. Brüel-Gabrielsson, **M. Yurochkin**, and J. Solomon. Rewiring with Positional Encodings for Graph Neural Networks. *Transactions on Machine Learning Research (TMLR)*, 2023.

[9] L. Ngweta, S. Maity, A. Gittens, Y. Sun, and **M. Yurochkin**. Simple Disentanglement of Style and Content in Visual Representations. In *International Conference on Machine Learning (ICML)*, 2023.

[10] K. Greenewald, A. Gu, **M. Yurochkin**, J. Solomon, and E. Chien. k-Mixup regularization for deep learning via optimal transport. *Transactions on Machine Learning Research (TMLR)*, 2023.

[11] S. Maity, **M. Yurochkin**, M. Banerjee, and Y. Sun. Understanding new tasks through the lens of training data via exponential tilting. In *International Conference on Learning Representations (ICLR)*, 2023.

[12] L. Li, N. Aigerman, V. Kim, J. Li, K. Greenewald, **M. Yurochkin**, and J. Solomon. Learning Proximal Operators to Discover Multiple Optima. In *International Conference on Learning Representations (ICLR)*, 2023.

[13] L. Li, Q. Liu, A. Korba, **M. Yurochkin**, and J. Solomon. Sampling with Mollified Interaction Energy Descent. In *International Conference on Learning Representations (ICLR)*, 2023.

[14] Z. Ashktorab, B. Hoover, M. Agarwal, C. Dugan, W. Geyer, H. B. Yang, and **M. Yurochkin**. Fairness Evaluation in Text Classification: Machine Learning Practitioner Perspectives of Individual and Group Fairness. In *CHI Conference on Human Factors in Computing Systems*, 2023.

[15] S. Xue, Y. Sun, and **M. Yurochkin**. Calibrated Data-Dependent Constraints with Exact Satisfaction Guarantees. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. **Oral**.

[16] D. Mukherjee, F. Petersen, **M. Yurochkin**, and Y. Sun. Domain Adaptation meets Individual Fairness. And they get along. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

[17] **M. Yurochkin** and Y. Sun. Communication-Efficient Model Fusion. In H. Ludwig and N. Baracaldo, editors, *Federated Learning: A Comprehensive Overview of Methods and Applications*, pages 145–176. Springer International Publishing, Cham, 2022.

[18] M. Agarwal, **M. Yurochkin**, and Y. Sun. Personalization in Federated Learning. In H. Ludwig and N. Baracaldo, editors, *Federated Learning: A Comprehensive Overview of Methods and Applications*, pages 71–98. Springer International Publishing, Cham, 2022.

[19] T. Shnitzer, **M. Yurochkin**, K. Greenewald, and J. Solomon. Log-Euclidean Signatures for Intrinsic Distances Between Unaligned Datasets. In *International Conference on Machine Learning (ICML)*, 2022.

[20] B.C. Kwon, U. Kartoun, S. Khurshid, **M. Yurochkin**, S. Maity, D. Brockman, A. Khera, P. Ellinor, S. Lubitz, and K. Ng. RMExplorer: A Visual Analytics Approach to Explore the Performance and the Fairness of Disease Risk Models on Population Subgroups. In *IEEE Visualization Conference (VIS) Short Papers*, 2022.

[21] I. Baldini, D. Wei, K. Ramamurthy, **M. Yurochkin**, and M. Singh. Your fairness may vary: Pretrained language model fairness in toxic text classification. In *Findings of ACL*, 2022.

[22] W. Stephenson, S. Ghosh, T. Nguyen, **M. Yurochkin**, S. Deshpande, and Broderick T. Measuring the sensitivity of Gaussian processes to kernel choice. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

[23] F. Petersen, D. Mukherjee, Y. Sun, and **M. Yurochkin**. Post-processing for Individual Fairness. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[24] M. Agarwal, **M. Yurochkin**, and Y. Sun. On sensitivity of meta-learning to support data. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[25] S. Maity, D. Mukherjee, **M. Yurochkin**, and Y. Sun. Does enforcing fairness mitigate biases caused by subpopulation shift? In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[26] V. Huynh, N. Ho, N. Dam, X. Nguyen, **M. Yurochkin**, H. Bui, and D. Phung. On efficient multilevel Clustering via Wasserstein distances. *Journal of Machine Learning Research (JMLR)*, 2021.

[27] D. Mukherjee, A. Guha, J. Solomon, Y. Sun, and **M. Yurochkin**. Outlier-Robust Optimal Transport. In *International Conference on Machine Learning (ICML)*, 2021.

[28] S. Maity, S. Xue, **M. Yurochkin**, and Y. Sun. Statistical inference for individual fairness. In *International Conference on Learning Representations (ICLR)*, 2021.

[29] A. Bower, H. Eftekhari, **M. Yurochkin**, and Y. Sun. Individually Fair Rankings. In *International Conference on Learning Representations (ICLR)*, 2021.

[30] A. Vargo, F. Zhang, **M. Yurochkin**, and Y. Sun. Individually Fair Gradient Boosting. In *International Conference on Learning Representations (ICLR)*, 2021. **Spotlight**.

[31] **M. Yurochkin** and Y. Sun. SenSeI: Sensitive Set Invariance for Enforcing Individual Fairness. In *International Conference on Learning Representations (ICLR)*, 2021. **Oral**.

[32] M. Weber, **M. Yurochkin**, S. Botros, and V. Markov. Black Loans Matter: Distributionally Robust Fairness for Fighting Subgroup Discrimination. *NeurIPS Fair AI in Finance Workshop*, 2020. **Spotlight Talk**.

[33] L. Li, A. Genevay, **M. Yurochkin**, and J. Solomon. Continuous Regularized Wasserstein Barycenters. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[34] S. Claici*, **M. Yurochkin***, S. Ghosh, and J. Solomon. Model Fusion with Kullback–Leibler Divergence. In *International Conference on Machine Learning (ICML)*, 2020.

[35] D. Mukherjee*, **M. Yurochkin***, M. Banerjee, and Y. Sun. Two Simple Ways to Learn Individual Fairness Metric from Data. In *International Conference on Machine Learning (ICML)*, 2020.

[36] S. Xue, **M. Yurochkin**, and Y. Sun. Auditing ML models for individual bias and unfairness. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

[37] H. Wang, **M. Yurochkin**, Y. Sun, D. Papailiopoulos, and Y. Khazaeni. Federated Learning with Matched Averaging. In *International Conference on Learning Representations (ICLR)*, 2020. **Oral**.

[38] **M. Yurochkin**, A. Bower, and Y. Sun. Training individually fair ML models with sensitive subspace robustness. In *International Conference on Learning Representations (ICLR)*, 2020. **Spotlight**.

[39] **M. Yurochkin**, S. Claici, E. Chien, F. Mirzazadeh, and J. Solomon. Hierarchical Optimal Transport for Document Representation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[40] **M. Yurochkin**, M. Agarwal, S. Ghosh, K. Greenewald, and N. Hoang. Statistical Model Aggregation via Parameter Matching. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[41] P. Monteiller, S. Claici, E. Chien, F. Mirzazadeh, J. Solomon, and **M. Yurochkin**. Alleviating Label Switching with Optimal Transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[42] **M. Yurochkin**, Z. Fan, A. Guha, P. Koutris, and X. Nguyen. Scalable inference of topic evolution via models for latent geometric structures. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[43] **M. Yurochkin***, A. Guha*, Y. Sun, and X. Nguyen. Dirichlet Simplex Nest and Geometric Inference. In *International Conference on Machine Learning (ICML)*, 2019. **Long Talk**.

[44] **M. Yurochkin**, M. Agarwal, S. Ghosh, K. Greenewald, N. Hoang, and Y. Khazaeni. Bayesian Nonparametric Federated Learning of Neural Networks. In *International Conference on Machine Learning (ICML)*, 2019.

[45] **M. Yurochkin**, S. Upadhyay, D. Bouneffouf, M. Agarwal, and Y. Khazaeni. Online Semi-Supervised Learning with Bandit Feedback. *ICLR Limited Labeled Data (LLD) Workshop*, 2019.

[46] **Mikhail Yurochkin**. *Geometric Inference in Bayesian Hierarchical Models with Applications to Topic Modeling*. PhD thesis, University of Michigan, 2018.

[47] **M. Yurochkin**, A. Guha, and X. Nguyen. Conic Scan-and-Cover algorithms for nonparametric topic modeling. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[48] **M. Yurochkin**, X. Nguyen, and N. Vasiloglou. Multi-way Interacting Regression via Factorization Machines. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[49] N. Ho, X. Nguyen, **M. Yurochkin**, H. Bui, V. Huynh, and D. Phung. Multilevel Clustering via Wasserstein Means. In *International Conference on Machine Learning (ICML)*, 2017.

[50] **M. Yurochkin** and X. Nguyen. Geometric Dirichlet Means algorithm for topic inference. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016.

Patents

- K. Greenewald, L. Jung, J. Solomon, **M. Yurochkin**, and Y. Zeng. A method for outlier robust subgroup inference via clustering in the gradient space. Filed on April 29, 2023.
- **M. Yurochkin**, D. Mukherjee, M. Banerjee, Y. Sun, and S. Upadhyay. Learning Mahalanobis Distance Metrics from Data. Filed on June 11, 2021.
- S. Upadhyay, **M. Yurochkin**, D. Mukherjee, Y. Sun, A. Bower, H. Eftekhari, A. Vargo, and F. Zhang. Training Individually Fair Machine Learning Algorithms via Distributionally Robust Optimization. Filed on March 25, 2021.
- Y. Khazaeni, E. Daly, C. Muise, and **M. Yurochkin**. Artificial Intelligence for Learning Path Recommendations. Filed on December 31, 2020.
- K. Greenewald, **M. Yurochkin**, M. Agarwal, S. Ghosh, N. Hoang and Y. Khazaeni. A method for combining pre-trained neural networks into a memory and computation efficient global model. Filed on September 20, 2019.
- S. Upadhyay, **M. Yurochkin**, M. Agarwal, D. Bouneffouf and Y. Khazaeni. Method for Online Partially Rewarded Learning. Filed on August 28, 2019.

Open-Source Software

- tinyBenchmarks, data and Python package for fast evaluation of LLMs: `huggingface.co/tinyBenchmarks`
- inFairness, Python package for Individual Fairness: `github.com/IBM/inFairness`
- Playground for comparing Group and Individual Fairness of toxic text detectors: temporarily behind VPN

Talks

Tutorials

- AI Fairness through Robustness @ AAAI 2023
- Fairness of Machine Learning in Search Engines @ CIKM 2022
- AI Fairness @ MIT-IBM THINK Industry Day 2022

Invited Talks

- Evaluating and Routing LLMs efficiently with Benchmarks
  @ Wells Fargo Quant Technical Seminar 2024
- Operationalizing Individual Fairness
  @ One World YoungStatS Webinar series 2023
- AI Fairness & Language Models
  @ Northern Germany delegation visit to the MIT-IBM Lab 2023
- Practical Individual Fairness
  @ TrustML Young Scientist Seminars 2022
- Black Loans Matter: Fighting Bias for AI Fairness in Lending
  @ KDD Workshop on Machine Learning in Finance (Keynote) 2021
- Model fusion via single-round FL
  @ Enterprise-Strength Federated Learning, ICML Expo 2021
- Practical Individual Fairness algorithms
  @ Toronto Machine Learning Summit 2021
  @ ODSC West 2021
  @ Foundations of Algorithmic Fairness (FAF) workshop 2021
  @ Research Talks for Enel 2021
  @ IBM Research, Cambridge Lab All-hands 2021
- Fusion of Heterogeneous Models in Federated Learning

@ The 2nd Annual Federated & Distributed/ Decentralized Machine Learning Conference 2021
- Federated Learning: Practice and Modern Algorithms
  @ ODSC Europe 2021
- Invited panelist
  @ Federated Learning, AI and Data Security Summit, HUB Security event 2021
- Fairness in mortgage lending
  @ Tech for Racial and Social Justice, UW-Madison Data Science Bazaar 2021
- Individual Fairness through Robustness
  @ Algorithmic fairness with statistical guarantees, CMStatistics session 2020
  @ Global Data Science Community meet-up, UniCredit event 2020
- Fairness in AI
  @ What's Next in AI: AI we can trust, MIT-IBM conference 2020
- Bayesian nonparametric fusion of heterogeneous models
  @ IBM Research Cambridge seminar series 2019
- Geometric Inference in Bayesian Hierarchical Models with Applications to Topic Modeling
  @ MIT Geometric Data Processing Group seminar 2018

Contributed Presentations

- Data Science Research Forum (poster) 2017
- Joint Statistical Meetings (speed session talk and poster) 2017
- Michigan Student Symposium for Interdisciplinary Statistical Sciences (talk) 2017
- Michigan Institute for Computational Discovery and Engineering Symposium (poster) 2016
- Michigan Student Symposium for Interdisciplinary Statistical Sciences (poster) 2016
- From Industrial Statistics to Data Science (poster) 2015

Awards

- Outstanding Technical Achievement Award
  for contributions to Federated Learning Security and Privacy (IBM internal) 2023
- Outstanding (O-level) Accomplishment (IBM internal) 2022
- IBM Creator: Courageous Leader (IBM internal) 2022
- Outstanding Technical Achievement Award (IBM internal) 2021
- EB1 Green Card (outstanding professor or researcher) 2021
- Research (A-level) Accomplishment (IBM internal) 2020
- Faces of IBM: AI Pioneer (IBM internal) 2020
- Outstanding Demonstration - Honorable Mention Award NeurIPS 2020
- Reviewer Award NeurIPS 2018, 2019
- Rackham Conference Travel Grant 2016, 2017
- Abramov Scholarship (awarded to less than 5% students at MIPT) 2010

Mentorship

- Ronald Xu, MIT Master's of Engineering (MEng) student 2024-Present
- Anthony Ou, MIT Master's of Engineering (MEng) student 2023-2024
- Break Through Tech AI (BTAI) @ MIT, Project Advisor for two student teams 2023
- Michael Feffer, IBM Research Tech for Justice summer PhD student intern 2023
- Hao Bang (Kevin) Yang, MIT Master's of Engineering (MEng) student 2022-2023
- Break Through Tech AI (BTAI) @ MIT, Project Advisor for a group of six students 2022
- Yuchen Zeng, MIT-IBM summer PhD student intern 2022
- Subha Maity, IBM Research Tech for Justice summer PhD student intern 2022
- Lilian Ngweta, IBM Research summer PhD student extern 2022, 2023

- Todd Zhou, Research Science Institute (RSI) high school student 2022
- Luann Jung, MIT Master's of Engineering (MEng) student (with J. Solomon) 2021-2022
- Yining Chen, MIT-IBM summer PhD student intern 2021
- Haingoharijao Fanririniaina Ramandiamanana,
  African Master's in Machine Intelligence (AMMI) student (with J. Solomon and D. Palmer) 2020
- Debarghya Mukherjee, MIT-IBM summer PhD student intern 2020
- Hongyi Wang, IBM Research summer PhD student intern 2019

Professional Activities

Organization and Service

- Practical Bayesian Methods for Big Data and Big Models workshop (co-organizer)
  @ MIT-IBM AI week 2019
- Optimal Transport seminar (with Y. Mroueh)
  @ IBM Research AI seminar series 2019

Reviewer

- Neural Information Processing Systems (NeurIPS) 2016-2022
- Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 2022
- International Conference on Learning Representations (ICLR) 2018-2021
- International Conference on Machine Learning (ICML) 2017-2019
- Bayesian Analysis (BA) 2019
- Journal of Machine Learning Research (JMLR) 2019
- Journal of Computational and Graphical Statistics (JCGS) 2016

Media coverage

- Interview for Expansion, one of the top media outlets in Mexico (in Spanish) 2023
- "AWANI Tonight: Creating a fairer AI", Malaysian news (1.75M subscribers) 2023
- "Debugging foundation models for bias", IBM Research 2022
- "New research helps make AI fairer in decision-making", IBM Research 2021
- "Finding a good read among billions of choices", MIT News 2019
- "Optimal Transport for Label Switching: Using Geometry to Solve Problems in AI", IBM Research 2019

Additional information

- Personal webpage: `moonfolk.github.io`
- LinkedIn: `www.linkedin.com/in/mikhail-yurochkin-a45659114`
- GitHub: `github.com/moonfolk`
- Google Scholar: `scholar.google.com/citations?user=QjBF9sUAAAAJ`