

Felipe Maia Polo

maiapolo@umich.edu | felipemaiapolo.github.io | github.com/felipemaiapolo

EDUCATION

University of Michigan <i>PhD in Statistics (Advised by Yuekai Sun and Moulinath Banerjee; GPA=3.96/4.00)</i>	Ann Arbor, USA Sep. 2021 – Aug. 2026
University of São Paulo <i>MSc in Statistics (Advised by Renato Vicente; GPA=10.00/10.00)</i>	São Paulo, Brazil Jan. 2019 – Oct. 2021
University of São Paulo <i>BSc in Economics (Advised by Daniel dos Santos; Best student award; GPA=8.80/10.00)</i>	Ribeirão Preto, Brazil Jan. 2014 – Dec. 2018
University of Tokyo <i>Exchange (Merit-based scholarship from the University of São Paulo; GPA=10.00/10.00)</i>	Tokyo, Japan Sep. 2017 – Mar. 2018

WORK EXPERIENCE

Institute of Foundation Models (MBZUAI) <i>Senior Research Scientist</i>	Sep. 2026 – Present Sunnyvale, USA
• I work on IFM's data team, focused on generating synthetic data for Large Language Model training. • Concretely, I develop data pipelines, where I turn tasks into real examples where models learn from.	
Google DeepMind <i>Student researcher</i>	Aug. 2025 – Dec. 2025 London, UK
• Worked with Isabela Albuquerque at Google DeepMind to develop efficient AI model evaluation methods. • Developed a statistical framework that cut human-annotation cost by $10\times$–$20\times$ while preserving precise, fine-grained measurement of model capabilities. Resulted in the paper “Rich Insights from Cheap Signals: Efficient Evaluations via Tensor Factorization.”	
MIT-IBM Watson AI Lab <i>AI Research Summer Intern</i>	May. 2025 – Aug. 2025 Cambridge, USA
• Worked on language model evaluation via game-based settings with Leshem Choshen, Kristjan Greenewald, and the MIT-IBM Watson AI Lab team. • Developed a statistical framework to infer LLM skills from dynamic game data (e.g., chess). Resulted in the paper “A Statistical Framework for Game-Based AI Evaluation” presented at a NeurIPS 25 workshop.	
MIT-IBM Watson AI Lab <i>AI Research Summer Intern</i>	May. 2024 – Aug. 2024 Cambridge, USA
• Spent the summer working on language model scaling laws with Mikhail Yurochkin and the MIT-IBM Watson AI Lab team. • Proposed a scaling-law model for LLM skills (e.g., reasoning) to predict multi-benchmark performance across model families. Resulted in the NeurIPS 2025 paper “Sloth: Scaling Laws for LLM Skills to Predict Multi-Benchmark Performance Across Families”.	
Department of Statistics of the University of Michigan <i>Research Assistant</i>	Sep. 2023 – Present Ann Arbor, USA
• This research position is funded by my advisors, Yuekai Sun and Moulinath Banerjee. We develop together new methods in statistics and machine learning.	
Advanced Institute for Artificial Intelligence (AI2) <i>Research Assistant</i>	Mar. 2020 – Aug. 2021 São Paulo, Brazil
• Worked closely with the LatAm Experian DataLab data scientists, developing research related to credit scoring.	
Tikal Tech <i>Data Science Consultant</i>	Jan. 2019 – Aug. 2021 São Paulo, Brazil
• Developed machine learning and natural language processing tools to be applied in the legal field.	

Neuron - Data Science and Artificial Intelligence

Jan. 2018 – Jan. 2019

Founder and former president

São Paulo, Brazil

- Founded and led the group to be a reference as a bridge between Academia and Industry. We organized seminars and workshops with researchers and tech entrepreneurs and produced data science materials in Portuguese.
- “Data Girls”, one of the most important data science initiatives targeted to women in Brazil, is a spin-off of this project.

Social Economics Lab (LEPES)

Mar. 2014–Jan. 2017

Research Assistant

Ribeirão Preto, Brazil

- Performed statistical analysis using statistical software (R and Stata) and produced reports for clients, a great part of them linked to public policy.

RESEARCH

Published and accepted papers

- Cai, Peiyao, Chengyu Cui, *Felipe Maia Polo*, Seamus Somerstep, Leshem Choshen, Mikhail Yurochkin, Moulinath Banerjee, Yuekai Sun, Kean Ming Tan, and Gongjun Xu. “A Latent Variable Framework for Scaling Laws in Large Language Models.” *Annals of Applied Statistics* (2027).
- Somerstep, Seamus, *Felipe Maia Polo*, Allysson Flavio Melo de Oliveira, Prattyush Mangal, Mírian Silva, Onkar Bhardwaj, Mikhail Yurochkin, and Subha Maity. “CARROT: A Cost-Aware Rate-Optimal Router.” *Annals of Applied Statistics* (2027).
- *Maia Polo, Felipe*, Aida Nematzadeh, Virginia Aglietti, Adam Fisch, and Isabela Albuquerque. “Rich Insights from Cheap Signals: Efficient Evaluations via Tensor Factorization.” *NeurIPS* (2026).
- *Maia Polo, Felipe*, Xinhe Wang, Mikhail Yurochkin, Gongjun Xu, Moulinath Banerjee, and Yuekai Sun. “Bridging Human and LLM Judgments: Understanding and Narrowing the Gap.” *NeurIPS* (2025).
- *Maia Polo, Felipe*, Seamus Somerstep, Leshem Choshen, Yuekai Sun, and Mikhail Yurochkin. “Sloth: scaling laws for LLM skills to predict multi-benchmark performance across families.” *NeurIPS* (2025).
- Züfle, Maïke, Vilém Zouhar, Tu Anh Dinh, *Felipe Maia Polo*, Jan Niehues, and Mrinmaya Sachan. “COMET-poly: Machine Translation Metric Grounded in Other Candidates.” *WMT2025* (2025).
- Somerstep, Seamus, *Felipe Maia Polo*, Moulinath Banerjee, Ya’acov Ritov, Mikhail Yurochkin, and Yuekai Sun. “A statistical framework for weak-to-strong generalization.” *ICLR* (2025).
- Shabtay, Nimrod, *Felipe Maia Polo*, Sivan Doveh, Wei Lin, M. Jehanzeb Mirza, Leshem Choshen, Mikhail Yurochkin et al. “LiveXiv—A Multi-Modal Live Benchmark Based on Arxiv Papers Content.” *ICLR* (2025).
- Bracale, Daniele, Subha Maity, *Felipe Maia Polo*, Seamus Somerstep, Moulinath Banerjee, and Yuekai Sun. “Microfoundation Inference for Strategic Prediction.” *AISTATS* (2025).
- *Maia Polo, Felipe*, Ronald Xu, Lucas Weber, Mírian Silva, Onkar Bhardwaj, Leshem Choshen, Allysson Flavio Melo de Oliveira, Yuekai Sun, and Mikhail Yurochkin. “Efficient multi-prompt evaluation of LLMs.” *NeurIPS* (2024).
- *Maia Polo, Felipe*, Subha Maity, Mikhail Yurochkin, Moulinath Banerjee, and Yuekai Sun. “Weak Supervision Performance Evaluation via Partial Identification.” *NeurIPS* (2024).
- *Maia Polo, Felipe*, Lucas Weber, Leshem Choshen, Yuekai Sun, Gongjun Xu, and Mikhail Yurochkin. “tinyBenchmarks: evaluating LLMs with fewer examples.” *ICML* (2024).
- Wang, Hongyi, *Felipe Maia Polo*, Yuekai Sun, Souvik Kundu, Eric Xing, and Mikhail Yurochkin. “Fusing Models with Complementary Expertise.” *ICLR* (2024).
- *Maia Polo, Felipe*, Yuekai Sun, and Moulinath Banerjee. “Conditional independence testing under misspecified inductive biases.” *NeurIPS* (2023). (**spotlight presentation**)
- *Maia Polo, Felipe*, Rafael Izbicki, Evanildo Gomes Lacerda Jr, Juan Pablo Ibieta-Jimenez, and Renato Vicente. “A unified framework for dataset shift diagnostics.” *Information Sciences* (2023).
- *Maia Polo, Felipe*, Stuti Saxena, and Charalampos Alexopoulos. “Stop the war, for Godsake!: A text analysis of speeches of Volodymyr Oleksandrovyh Zelenskyy, the Ukrainian President.” *International Journal of Public Administration* (2023).
- *Maia Polo, Felipe*, and Renato Vicente. “Effective sample size, dimensionality, and generalization in covariate shift adaptation.” *Neural Computing and Applications* (2022): 1-13.
- *Maia Polo, Felipe*, et al. “LegalNLP-Natural Language Processing methods for the Brazilian Legal Language.” *Anais do XVIII Encontro Nacional de Inteligência Artificial e Computacional*. SBC (2021).
- *Maia Polo, Felipe*, Itamar Ciochetti, and Emerson Bertolo. “Predicting legal proceedings status: approaches based on sequential text data.” *Eighteenth International Conference on Artificial Intelligence and Law* (2021).

Preprints

- *Maia Polo, Felipe*, Leshem Choshen, Yuekai Sun, and Kristjan Greenewald. "A Statistical Framework for Game-Based AI Evaluation." In NeurIPS 2025 Workshop on Evaluating the Evolving LLM Lifecycle: Benchmarks, Emergent Abilities, and Scaling (2025).
- Patron, Gabriel, Zhiwei Xu, Ishan Kapnadak, and *Felipe Maia Polo*. "Recommendations Beyond Catalogs: Diffusion Models for Personalized Generation." arXiv preprint arXiv:2502.18477 (2025).

Book chapters

- *Maia Polo, Felipe*, and Daniel Domingues dos Santos. "Educação Infantil: Avanços e desafios para o futuro próximo." in Lent, Robert, Augusto Buchweitz, and Mailce Borges Mota [organizers]. "Ciência para Educação: Uma ponte entre dois mundos." São Paulo: Editora Atheneu (2018): 73-96.
 - * In 2019, the book was a finalist in the most important Brazilian prize for books, the "Jabuti" prize.

AWARDS

Academic

- Rackham Predoctoral Fellowship from the University of Michigan (2025-2026). This fellowship supports students working on dissertations that are unusually creative, ambitious, and impactful.
- Outstanding Reviewer. Journal of Educational and Behavioral Statistics (JEBS) and the American Educational Research Association (AERA) (2025).
- NeurIPS 2025 Top Reviewer (top 8%). Complimentary registration to the conference.
- Best presentation award. The 2025 Michigan Student Symposium for Interdisciplinary Statistical Sciences (2025).
- Best poster presentation. Department of Statistics of the University of Michigan (2023).
- Honorable mention of the poster session. The 2023 Michigan Student Symposium for Interdisciplinary Statistical Sciences (2023).
- NeurIPS Scholar Award (travel grant) (2023).
- Best Economics Student. University of São Paulo (2018).
- Best Economics Students in the State of São Paulo. Regional Economics Council of State of São Paulo (2018)

Hackathons and other competitions

- Runner-up in *NLP Hackathon Dotz* (2020). Prize: US\$3000 credit in Google Cloud + R\$2000 (Equivalent to 2 monthly minimum wages at that time)
- Champion team in the Economics Olympiad of São Paulo (2018). Prize: R\$4,000 (Equivalent to 4 monthly minimum wages at that time).
- Champion Team of "Desafio Claro" @ Hackathon Serasa E3Hack Experiance JAM5 (2018). Prize: R\$10,000 (Equivalent to 10 monthly minimum wages at that time).
- Participation in the Hult Prize Regional Finals (2018). The Hult Prize is a social business international championship. My team was selected among thousands worldwide to compete in the Mexico City Final.

PRESENTATIONS

Invited talks

- "Statistical approach to AI evaluation" at PatronusAI (2025).
- "Psychometrics for LLM evaluation" at the University of Sao Paulo (2025).
- "Psychometrics for LLM evaluation" at FGV-EMAp (2025).
- "Efficient evaluation of LLMs with tinyBenchmarks and PromptEval." at ServiceNow (2024).
- "Efficient evaluation of LLMs with tinyBenchmarks and PromptEval." at Google DeepMind (2024).
- "Efficient multi-prompt evaluation of LLMs." at Unify (2024).
- "tinyBenchmarks: evaluating LLMs with fewer examples." at Google DeepMind (2024).

SERVICE

Reviewing

- NeurIPS 2024, ICLR 2025, ICML 2025, NeurIPS 2025, NeurIPS 2026, Journal of Educational and Behavioral Statistics (JEBS) 2025.

Committees

- Diversity, equity, and inclusion committee (Winter 2023): Participated in recurrent meetings to discuss DEI topics with leaders of the Department of Statistics at the University of Michigan.
- Students' Representative Council (2016): Represented the Economics major students at the University of São Paulo in meetings with professors and leaders of the department during the whole year.

Events

- The Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSIIS) (2024): Led a team of eight graduate students for MSSIIS 2024 organization. Our team increased the number of abstract submissions by 30% relatively to past years.
- Statistics Student Seminar (Fall 2023): Responsible for organizing weekly seminars targeted to graduate students of the Department of Statistics at the University of Michigan.
- Data Science Talks Meetup (2019): Led the organization of two editions of the data science meetup "Data Science Talks", a unique initiative at the University of São Paulo.
- Data Science Day (2018): Led the organization of the first edition of "Data Science Day", one of the biggest Data Science events in São Paulo. Partnered with companies such as Microsoft, Accenture, Movel, and Udacity to deliver speeches or participate in the panels.

TEACHING

University of Michigan

- Data Mining and Statistical Learning (Stats 415, Undergraduate Level - Winter 2023).
- Introduction to Statistics (Stats 250, Undergraduate Level - Winter 2022/Fall 2022)
- Regression Analysis (Stats 500, Graduate Level - Fall 2021).

University of São Paulo

- Introduction to Probability and Statistics II (Undergraduate Level - 2019).
- Introduction to Statistical Inference (Undergraduate Level - 2018).
- Introduction to Probability and Statistics I (Undergraduate Level - 2017).
- Integral and Differential Calculus II (Undergraduate Level - 2015).
- Integral and Differential Calculus I (Undergraduate Level - 2015).