

# Michael S. Truong

 [emstruong.github.io](https://github.com/emstruong) |  [github.com/emstruong](https://github.com/emstruong) |  [linkedin.com/in/emstruong](https://www.linkedin.com/in/emstruong) |  [emstruonger@gmail.com](mailto:emstruonger@gmail.com)

Quantitative methods Ph.D. candidate (expected June 2027) who builds and validates statistical machine learning methods: tree-based structural equation modeling, large-scale Monte Carlo simulation on high-performance computing clusters, and simulation-based checks of LLM-generated statistical code. Seeking full-time data science, data engineering, or machine learning research roles in Canada. Canadian citizen based in Toronto, Ontario.

## EDUCATION

---

**Ph.D. in Psychology (*Quantitative Methods*)** Expected June 2027

York University (Supervisor: *Dr. David B. Flora*)

Dissertation: *New Avenues in Machine Learning Structural Equation Modeling: IGSCA Trees*

**M.A. in Psychology (*Developmental Sciences*)** 2022

York University (Supervisors: *Drs. Thanujeni Pathman & W. Dale Stevens*)

**H.B.Sc. in Neuroscience & Psychology** 2020

University of Toronto (Supervisors: *Dr. Morris Moscovitch & Nichole Bouffard*)

## TECHNICAL SKILLS

---

**Programming:** R (expert; package development, testthat), Stan, Quarto, Bash; Python and SQL (developing)

**Statistics & Machine Learning:** Tree-Based Machine Learning, Monte Carlo Simulation, Causal Inference, Bayesian Modeling, Structural Equation Modeling, Mixed-Effects Modeling, Categorical Data Analysis, Network Psychometrics

**Data Engineering & HPC:** Git, Singularity/Docker, Slurm (Digital Research Alliance of Canada), Arrow, DuckDB, targets pipelines, Linux

**AI-Assisted Development:** Claude Code; validating LLM-generated statistical code with Monte Carlo simulations

## PROFESSIONAL & RESEARCH EXPERIENCE

---

**Doctoral Researcher** 2022 - Present

York University (Supervisor: *Dr. David B. Flora*)

- Designed IGSCA Trees, a tree-based machine learning method that detects subgroups with different structural equation model parameters (dissertation; paper under review)
- Validated the method with large-scale Monte Carlo simulation studies, in collaboration with Drs. Florian Schuberth and Ji Yeh Choi
- Adapted principles of simulation-based calibration to check the correctness of LLM-generated statistical software
- Compared frequentist and Bayesian multilevel models when cluster size is a confounder (paper with Dr. Xijuan (Cathy) Zhang), running three simulation studies that used 450+ CPU core-years on Digital Research Alliance of Canada clusters
- Presented at the International Meeting of the Psychometric Society (2025, 2026) and Modern Modeling Methods (2024)

**Open-Source R Package Contributor** 2023 - Present

[github.com/emstruong](https://github.com/emstruong)

- [cSEM R Package](#)
  - Implemented IGSCA Trees and integrated GSCA-type models with the package's shared infrastructure
  - Reduced runtime of GSCA-type estimation routines by 50% while improving numerical stability

- Updated testthat test infrastructure
- [SimDesign R Package](#)
  - Quality assurance on package features via GitHub bug reports (credited in [package NEWS](#))
  - Filed several successful feature requests
  - Implementing improved handling of parallelized workloads via load balancing
- [blavaan R Package](#)
  - Facilitated reporting of model fits with new tidy *broom* methods
    - Initially drafted using Claude Opus and manually improved to match coding standards
  - Improved data analytical UX by adding brms-style default arguments
- [lavaan R Package](#)
  - Reported a bug in the handling of likelihood-ratio tests
  - Improved UX of fitting multigroup models
- [glmmTMB R Package](#)
  - Identified cross-platform differences in handling of model convergence

## Statistical Consultant

2023 - 2026

York University, [Statistical Consulting Service](#)

- Advised graduate and faculty clients and gave weekly methods seminars, winning the Mirka Ondrack Award for contributions to statistical consulting

## Peer-Review and Open Science

2022 - Present

- Reviewed manuscripts submitted to Q1 quantitative methods journals, including [Advances in Methods and Practices in Psychological Science](#), for novelty, correctness, and fit
- Reviewed a public [dataset](#) for fidelity and worked with authors to obtain resolutions

## Teaching Assistant and Guest Lecturer

2020 - Present

York University

- Practicum for Statistical Consulting (MATH6627); Univariate Analysis II: Regression (PSYC6132)

## Research Assistant

2024 - 2025

York University (Supervisor: *Dr. Michael Friendly*)

- Reviewed [Visualizing Multivariate Linear Models](#)

## SELECTED PUBLICATIONS

### Exploring the Multiverse with Decision Trees

Accepted, [AMPPS](#)

*Truong, M. S., Crone, G., & Choi, J. Y.*

### Robust Recursive SEM with IGSCA Trees

Under review

*Truong, M. S., Schuberth, F., Choi, J. Y., & Flora, D. B.*

### What if Cluster Size is the Confound in Multilevel Models?

In preparation

*Truong, M. S., Zhang, X., & Flora, D. B.*

## SELECTED AWARDS

2025 SSHRC Doctoral Fellowship

**\$80,000**

2024 Mirka Ondrack Award for Contributions to Statistical Consulting

2023 [Canadian Psychological Association: Quantitative Methods Student Presentation Award](#)