MINGXUAN (ADRIAN) CAO

(last update: October 4, 2025)
1000 South Clark Street, Chicago, IL 60605
(314) · 536 · 1645 ◇ caom@uchicago.edu ◇ https://adriancmx.github.io/

EDUCATION

University of Chicago *Sept. 2024 - Now*
Ph.D. in Data Science
Research Interest: Biological Generative Models; Reinforcement Learning; High-dimensional Statistics

Washington University in St. Louis *Sept. 2020 - May 2024*
AB+AM in Statistics (Honors)
Second Major in Economics and Strategy
Minor in History

RESEARCH INTEREST

My research builds **robust, inference-time guidance and alignment methods** at the interface of **statistical learning**, **diffusion/SDE generative models**, and **reinforcement learning**. I aim to enable purposeful scientific discovery in biology and chemistry by designing **scalable algorithms for planning, sampling, and multi-objective optimization** in complex sequence–structure domains.

PAPERS & REPORTS

M. Cao, Xuefeng Liu, Songhao Jiang, Xiao Luo, Xiaotian Duan, Mengdi Wang, Tobin R. Sosnick, Jinbo Xu, Rick L. Stevens "Monte Carlo Tree Diffusion with Multiple Experts." [arXiv] **In Submission.**

M. Cao, Xuefeng Liu, Songhao Jiang, Qinan Huang, Tinson Xu, Ian Foster, Mengdi Wang, Hening Lin,Rick L. Stevens " RL Reasoning with Molecular Design: An LLM-Agent Framework Orchestrating Specialized Foundation Models." **In Preparation.**

M. Cao, Chetkar Jha. "Estimating number of factors in factor analysis for datasets with missing entries at block level." **In Preparation.**

M. Cao, Honglin Bao, Jiawei Zhang, James A. Evans, "From Division to Unity: A Large-Scale Study on the Emergence of Computational Social Science, 1990-2021." **Proceedings of the ACM on Web Conference 2025 (WWW '25).** [arXiv] [Visualization]

THESIS

Bootstrap Estimation of Pre-averaged Volatility under Microstructure Noise. *Advised by Prof. José Figueroa-López.*

RESEARCH EXPERIENCE

Agentic AI for Biological Design and Inference June 2025 - Present
Advisor: Rick L. Stevens *University of Chicago / Argonne*

- Our goal is to build AI systems that *plan* and *reason* over biological objectives—designing molecules/proteins and inferring mechanisms—under realistic, multi-objective constraints (potency/ADMET/synthesizability; scaffold/motif retention; edit budgets). Methodologically, we combine an agentic LLM controller for decomposition and routing with reinforcement learning for decision-making, expert generators for scaffold/fragment/controllable edits, and structure-aware inference-time guidance. For proteins, we integrate *Monte Carlo Tree Diffusion* with pLDDT-guided masking and PH-UCT selection for efficient multi-token edits. Across antiviral/oncology molecular tasks and folding/inverse-folding protein benchmarks, we target improvements in validity, scaffold similarity, and task-specific design quality while developing general, test-time controllable tools for fine-grained biological design and mechanistic inference.

Interpreting GWAS Signals with Protein/Regulatory Foundation Models Dec. 2024 - May 2025
Advisor: Hae Kyung Im and Dan L. Nicolae *University of Chicago*

- Developed interpretable pipelines to link GWAS loci to molecular mechanisms by combining protein language models and regulatory predictors: (i) used PLM embeddings/channels to quantify how coding SNPs perturb residue environments and functional motifs as pathogenicity surrogates; (ii) modeled noncoding variants with eQTL-informed effect directions and sequence-to-expression predictors (PrediXcan/TFXcan-style, Enformer-like scores) to connect allelic dosage to gene expression and downstream traits while explicitly acknowledging tissue/context specificity and power gaps underlying partial GWAS–eQTL overlap; and (iii) treated selection metrics (e.g., iHS, F_{ST}) as exploratory annotations only, with causal claims relying on LD-aware fine-mapping/colocalization—complemented by a preliminary theoretical result toward fine-mapping without external LD reference panels.

Deterministic Parallel Analysis under Missing at Block Level May 2023 – Present
Advisor: Chetkar Jha *Washington University*

- Extended parallel-analysis factor selection to MAR settings and extending to MNAR by incorporating a shifted Marčenko –Pastur law for eigenvalue thresholds; provided corrections for rank recovery in high dimensions with missingness, and empirically compared against likelihood and spectral baselines across SNR/missingness regimes, showing improved recovery at low ranks.

A Machine Learning Approach to Predicting Opioid Use in Women with Breast Cancer Apr. 2023 - May 2024
Advisor: Prajakta Masurkar *AMGEN Inc.*

- Applied regularized GLMs, tree ensembles, and calibrated risk scores to predict high-risk opioid use following cancer treatment under class imbalance; performed covariate-shift sensitivity analyses across comorbidities, conducted subgroup-wise evaluation with stratified metrics and decision-curve analysis to quantify intervention value, and explored SGD-based causal adjustments (propensity weighting/overlap weighting) to probe treatment-response heterogeneity.

High-dimensional Time Series Change Point Detection in Neuronal Activity Analysis Apr. 2023 - Aug. 2023
Advisor: Likai Chen *Washington University*

- Implemented moving-sum and moving-quantile detectors for multi-neuron spike-train time series; derived practical detection thresholds under non-Gaussian noise, built a subsectioning/selection pipeline over 65 simultaneously recorded neurons to localize behavioral state transitions (sleep/wake focus), and benchmarked latency–false-alarm trade-offs via permutation and block-bootstrap resampling.

Bootstrap Estimation of Pre-averaged Volatility under Microstructure Noise May 2022 – Apr. 2023
Advisor: José Figueroa-López *Washington University*

- Studied finite-sample distortions of pre-averaging–based volatility estimators in diffusions with market microstructure noise; proposed an online bootstrap combining wild and block resampling for dependent data to improve coverage and reduce bias in high-frequency settings.

PROFESSIONAL EXPERIENCE

MorningStar, Inc. June 2023 – Aug. 2023
Investor Relations Intern *Chicago, IL*

- Performed comprehensive data analysis, including competitor analysis, examining growth, profitability, and other key metrics, utilizing data visualization to communicate information in quarterly earnings release and presentation and board meeting.
- Applied quantitative research skills, including AI techniques, to extract insights from financial data, aiding in identifying market trends and providing knowledge of investment strategies, financial products, and industry dynamics for investors' questions.

Shanghai Seeking Sense Investment Management Co., Ltd July 2022 - Aug. 2022
Quantitative Investment Analyst *Shanghai, China*

- Developed industry rotation models using PSY unit root test and Bayesian Online Changepoint Detection to detect bubble signals in fundamentals, achieving a 140+% annualized return and 13% max retracement.
- Researched convertible-bond-based strategy and formed stable income model that generates 80+% annualized returns by seizing low premium rate and low-price bonds using R and Python; studied value appropriation of industrial chain of chips in China.

TEACHING & MENTORING

Teaching Assistant at University of Chicago

- DATA 30100: Intro to Data Science (TA) — Autumn 2025
- DATA 21300: Models in Data Science (TA) — Spring 2025

Teaching Assistant at Washington University

- Math 4211: Statistics for Data Science II (AI & Grader) — Spring 2024
- Math 461/5155: Time Series Analysis (AI & Grader) — Fall 2023
- Math 459: Bayesian Statistics (UTA & Grader) — Spring 2023
- Math 495: Stochastic Processes (UTA & Grader) — Spring 2023
- Math 456: Topics in Financial Mathematics (UTA & Grader) — Fall 2022
- Math 4211: Statistics for Data Science II (UTA & Grader) — Spring 2022

HONORS & AWARDS

- Magna Cum Laude, Washington University — 2023
- Highest Distinction Graduate in Statistics — 2023
- 2023 MCM/ICM Honorable Mention — 2023
- 2023 Freiwald Scholars Fellowship in the Department of Math and Statistics of Washington University — 2023

CONFERENCE & WORKSHOP

- Committee of Princeton Fintech & Quant Conference in Chicago — Oct. 2023
- Washington University Undergraduate Research Symposium — Apr. 2023

TECHNICAL STRENGTHS

Computer Languages	Pascal, C++, Java, R, Stata, Python, MATLAB, SQL, LaTex, Mathematica, SAS
Languages	English (bilingual), Mandarin Chinese (native), French (elementary)