Caleb Maresca

646-785-1968 • marescacc@gmail.com • linkedin.com/in/caleb-maresca

PhD student specializing in machine learning and mechanistic interpretability. Experienced in building end-to-end ML systems with deep learning architectures such as transformers and RNNs. Curious to understand how neural networks work on a fundamental level and passionate about developing more transparent and safe AI systems.

Experience

AI Safety Camp – Research Team Member Jan 2025 – Present

Working on developing novel safety mechanisms for reinforcement learning systems.

Oxford AI Safety Initiative – ARBOx Participant Jan 2025

Completed an intensive Machine Learning Safety bootcamp focused on practical implementation and research replication. Key accomplishments include:
- Won third place project for analyzing and proposing extensions to the novel Monet (Mixture of Monosemantic Experts) architecture
- Implemented GPT-2 small from scratch
- Developed and applied mechanistic interpretability techniques including sparse autoencoders
- Implemented modern reinforcement learning algorithms including Deep Q-Learning and PPO

New York University – Research Assistant Aug 2024 – Jan 2025

Contributed to research leveraging LSTM models to quantify investor bias.

New York University – Teaching Assistant Aug 2023 – Jan 2024

Developed and led tutorial sessions for Microeconomics 1 (PhD).

Hebrew University of Jerusalem – Research Assistant 2020 – 2022

Led statistical analyses and mathematical modeling for behavioral economics research with Professor Ori Heffetz, developing simulations and theoretical proofs for published work.

Education

New York University – PhD Student – Economics 2022 – Present

GPA: 3.738

Relevant coursework:

Foundations of Machine Learning and Deep Learning – Deep neural networks, RNNs, CNNs, attention and transformers, autoencoders, GANs

Deep Learning and LLM Systems – Distributed and cloud-based training, MLOps, LLM pre-training and fine-tuning, RAG and agents, efficient LLM serving, RLHF

Econometrics IV – Stochastic gradient descent, causal machine learning

Econometrics II – Time series models, ARMA, VARs

Math for Economists II – Measure theory and measure theoretic probability theory

Math for Economists I – Real analysis

Hebrew University of Jerusalem – MA – Economics 2021 – 2022

Cumulative average: 99%

Hebrew University of Jerusalem – BA – Economics 2018 – 2021

Cumulative average: 96%

Professional Skills

- Programming Languages: Python, R, Julia
- Machine Learning & Deep Learning: PyTorch, TensorFlow, Ray, WandB, Transformers, CNNs, RNNs
- Data Science: NumPy, Pandas, Scikit-learn
- Data Visualization: Matplotlib, Seaborn
- Statistical Methods: Time Series Analysis, Causal Inference, Econometrics
- Development Tools & Systems: Git, Linux

Projects

Multi-Stock Return Prediction Using LLMs and Financial News; with Nishant Asati

GitHub: https://github.com/CalebMaresca/nscan

Developed novel methodology integrating financial news analysis with stock embeddings using Differential Transformers to predict multiple stock returns simultaneously. Implemented end-to-end ML pipeline including data preprocessing, model architecture design, distributed training and hyperparameter optimization.

Double Machine Learning Method Implementation and Analysis

Replicated Chernozhukov et al. (2018) implementing double/debiased machine learning in Python. Implemented multiple ML methods with hyperparameter tuning and compared performance to DoubleML package, achieving faster runtime.

Working Papers

Racing for the Future: Capital Accumulation Before Transformative AI

Developed theoretical model analyzing how expectations of advanced AI affect current economic decisions. Shows how anticipated automation creates novel savings incentives as households compete for future control of AI labor.

Happier Than Thou, Causal Evidence for the Effect of Religion on Subjective Well-Being; coauthored with Joseph Lee

Applied Monotone Instrumental Variables (MIV) and Monotone Treatment Selection (MTS) to World Values Survey data to establish causal relationship between religious belief and subjective well-being.

The (In)Effectiveness of State R&D Grants

Applied synthetic control methods to evaluate impact of state R&D subsidies on private investment and business formation, establishing upper bounds on program effects.

Cognitive Biases are Critical in Conflict Bargaining

Integrated prospect theory into a game-theoretic bargaining model, demonstrating significant effects of cognitive biases on equilibrium outcomes.