

Peter Kirgis

105 Taylor Court, Princeton NJ 08540
516-270-5452 — pk7019@princeton.edu

Education

Princeton School of Public and International Affairs

Expected May 2025

Master of Public Affairs (MPA)

- Cumulative GPA: 3.92
- Graduate Certificate in Statistics and Machine Learning
- Archibald A. Gulick Fellow in Public and International Affairs
- Relevant Coursework: AI Alignment and Safety, Intro to Reinforcement Learning, Machine Learning and Pattern Recognition, Societal Impacts of Algorithms and AI, Ethical Challenges from Advances in Technology
- Invited to present capstone project at the 2025 PICSciE/CSSML Joint Colloquium

Capstone Project: *Differences in the Moral Foundations of Large Language Models*

Advisor: Professor Peter Henderson

Large language models are increasingly being used in critical domains of politics, business, and education, but the nature of their normative ethical judgment remains opaque. Alignment research has, to date, not sufficiently utilized perspectives and insights from the field of moral psychology to inform training and evaluation of frontier models. I perform a series of synthetic experiments on a wide range of models from most major model providers using Jonathan Haidt's influential moral foundations theory (MFT) to elicit diverse value judgments from LLMs. I then use principal component analysis (PCA) to validate the construct of MFT in my LLM sample and explain variation in elicited moral judgment across models. My results suggest that models display different moral foundations from one another and from a nationally representative human baseline. This work seeks to spur further analysis of LLMs using MFT, including finetuning of open-source models, and greater deliberation by policymakers on the importance of moral foundations for LLM alignment. [Github Repo](#)

Williams College

June 2020

Bachelor of Arts with Honors in Political Economy and Philosophy

- Cumulative GPA: 3.65; Major GPA Political Economy: 3.74
- Concentration in Cognitive Science
- Relevant Coursework: Econometrics, Multivariable Calculus

Research Experience

HAL - Holistic Agent Leaderboard

Spring 2025

Contributor

- Working with Benedikt Stroebl, Sayash Kapoor, and Arvind Narayanan to build a standardized, cost-aware, and third-party leaderboard for evaluating agents.
- Independently implemented multiple benchmarks, SciCode and Assistant Bench, into the HAL harness.
- Built agents optimized for each of these benchmarks for an upcoming paper.

Work Experience

United States Census Bureau

June 2024 – August 2024

Data Scientist Fellow

- Selected from an applicant pool of 4,500 with an acceptance rate of 2% as part of the Coding it Forward Fellowship, a program that empowers early-career technologists to explore technology careers in government.
- Built sorting algorithms and k-nearest neighbors regression models in Python to improve existing imputation methods within the Economic Statistical Methods Division.
- Designed data visualizations using Python to compare imputation strategies using several loss methods.
- Final Presentation: Improving Imputation Accuracy

Commonwealth of Massachusetts

November 2021 – July 2023

Data Analyst

- Reported to the Commonwealth's Chief Data Officer as part of a data sharing and program evaluation team.
- Led inter-agency meetings on data analysis and visualization design.
- Created and maintained ETL pipelines in PostgreSQL to feed live visualizations.
- Generated analytical reports for agencies using correlational and machine learning techniques.
- Played a key role in implementing the Commonwealth's first differential privacy algorithm on longitudinal datasets relating to early childhood education and workforce development programs.
- Built scripts to fully automate accessible data downloads for agency data stories using Python and AWS Lambda.
- Designed interactive data visualizations relating social safety net participation to wage growth using polynomial regression in Tableau.
- Presented summary of year's work to Boston Area Research Initiative at Fall 2022 meeting.

Honors and Awards

Equity in Governance Award – Commonwealth of Massachusetts

November 2022

- One of five annual performance recognition awards given out by governor's office.
- Given for contributions to the Commonwealth's first public-facing employee diversity dashboard (State Employee Diversity Dashboard).

2019 Williams College Venture Challenge – First Prize

April 2019

- Won \$15,000 as part of a team of three Williams students.
- Created a business proposal for an alternative method of concussion diagnosis using continuous biometric data.
- Conducted meetings with tech executives to discuss market strategies and product design.