

Technical Skills:

Python (Keras), R (dplyr), SQL, C, Java, Javascript, LaTeX, Django, Node.js

Yale Coursework:

STAT 365

Machine Learning

ECON 421a

Design of the Digital Economy

CPSC 477

Natural Language Processing

CPSC 458

Automated Decision Making

LING 227

Language and Computation

CPSC 323

Systems Programming

CPSC 662

Spectral Graph Theory

CPSC 437

Database Systems

CPSC 469

Randomized Algorithms

CPSC 365

Algorithms

MATH 305

Real Analysis

MGT 656

Management Of Software Development

CPSC 223

Data Structures and Programming

MATH 345

Modern Combinatorics

MATH 244

Discrete Mathematics

Awards:

Pennapps: Best Use of Spotify API

4x AIME Qualifier

William Cai

william.cai@yale.edu

About me I am a Research Assistant in the Computational Social Science group at Microsoft Research. I use data to improve societal systems.

Portfolios @ whereiswillcai.com, github.com/wicai.

Education

2013-2017, Yale University

- B.S. with distinction in Computer Science, 3.66 GPA (3.82 Major)

Experience

2017-Present, Research Assistant, Microsoft Research NYC

- Project Ratio: Platform to analyze all online news using machine learning. Named entity recognition and global resolution, clustering on daily news stories over thousands of sources to detect events.
- Created infrastructure to run surveys on various crowdsourcing platforms (Mechanical Turk, Pollfish) and performed data analysis for social psychology experiments.

2016, Software Engineering Intern, Google

- Designed and built pipeline to analyze the regression range of candidate flaky tests in the Chromium codebase. Automated swarming reruns over code commits at different points in time in order to determine when the tests became flaky.

2016-Present, Research Assistant, Yale Institute for Social and Policy Studies

- Applying network models with Prof. Papachristos to identify police officers who are more likely to discharge their firearm.

2015, Research Assistant, Yale Institute for Network Science

- Invented algorithm for linear regression over a distributed system, sending fractions of eigen-decompositions of the covariance matrix.

2014, Caltech SURF Fellow, NASA's Jet Propulsion Laboratory

- Estimated cosmic parameters with MCMC on computing cluster.

Projects

2016, City Design for Self Driving Cars, www.whereiswillcai.com/dde.html

- White paper describing potential problems and economic solutions related to the onset of self-driving vehicle technology. Simulation of the traffic of a city with varying penetrations of self-driving cars.

2015, Clusterfy, github.com/wicai/Music-Visualization

- Webapp in Flask that uses the Spotify and Echo Nest APIs to scrape a user's songs and associated data. Visualization with the Highcharts javascript library. "Best Use Of Spotify/EchoNest API" at Pennapps.