

# Zachary Novack

[zacharynovack.github.io](https://zacharynovack.github.io)

znovack@ucsd.edu

(561) 866-0646

**RESEARCH INTERESTS** Machine Learning (ML) for education, music information retrieval, empirical deep learning, computational social science

**EDUCATION BACKGROUND** *Ph.D. in Computer Science* Fall 2022 - May 2027 (Expected)  
[University of California, San Diego](#), San Diego, CA

- Advised by Prof. Julian McAuley
- *Selectec Coursework:*
  - Deep Generative Models, Search and Optimization, Information Visualization, Recommender Systems
  - *In Progress: Computing Education Research, Design Prototyping*

*B.S. in Statistics & Machine Learning* August 2018 - May 2022  
[Carnegie Mellon University](#), Pittsburgh, PA

- Minor in Sonic Arts (music technology)
- 3.93/4.0 GPA
- *Selected Coursework:*
  - Probabilistic Graphical Models, Convex Optimization, Multimedia Signal Processing, ML w/Large Datasets, Real Analysis, Numerical Linear Algebra, Probability & Statistics, Statistical Computing, Linear Algebra, Philosophy of ML, Algorithms & Data Structures

**RESEARCH EXPERIENCE** *Research Assistant* Spring 2021 - Present  
[ACMI Lab](#), Carnegie Mellon University, Pittsburgh, PA

- Investigated how to recycle the final layer of pre-trained neural networks, rather than retraining it from scratch, for few-shot classification tasks (ongoing).
- Proposed new method to leverage hierarchical class information for zero-shot prediction in CLIP models (in submission).
- Designed a large-scale verification study validate explicit regularization mechanisms for SGD across modern image benchmarks and model types (in submission).

*Research Assistant* Summer 2020 - Present  
[Laboratory for Social Minds](#), Carnegie Mellon University, Pittsburgh, PA

- Designed a Bayesian autoregressive model to analyze addiction effects on social media websites (ongoing).
- Implemented Latent Dirichlet Allocation (LDA) to investigate ideological network evolution on the fringe web forums /pol/ (4chan) and The Red Pill (Reddit).

*Undergraduate Researcher* Summer 2020 - Spring 2021  
[Dietrich College](#), Carnegie Mellon University, Pittsburgh, PA

- Constructed filtering algorithm to parse sparse text documents for specific topic occurrences
- Modified existing sentiment analysis implementation to account for valence-shifters in congressional speeches
- Implemented multiple behavioral game theoretic models in matlab to simulate strategic choice patterns in asymmetric two-player games

## PAPERS & PUBLIC WORKS

### *Conference Papers*

- **Disentangling the Mechanisms Behind Implicit Regularization in SGD**  
Zachary Novack, Simran Kaur, Tanya Marwah, Saurabh Garg, Zachary Lipton  
International Conference on Learning Representations (ICLR), 2023  
**Spotlight** and **Best Poster** at NeurIPS Workshop on The Benefits of Higher-Order Optimization in Machine Learning, 2022

### *Workshops / Preprints*

- **CHiLS: Zero-Shot Image Classification with Hierarchical Label Sets**  
Zachary Novack, Saurabh Garg, Zachary Lipton  
ICLR Workshop on Multimodal Representation Learning, 2023

### *Nonrefereed Papers*

- **Down the Rabbit Hole: Modeling Twitter Dynamics through Bayesian Inference**  
Zachary Novack  
Senior Honors Thesis (Carnegie Mellon University), 2022
- **Personalized Sequential Recommendation for Adaptive Itemization in MOBA Games**  
Zachary Novack  
Web Mining and Recommender Systems (CSE 258) Course Project (UC San Diego), 2022
- **Towards Generalizable Deep Speech Anonymization**  
Aaron Broukhim, Zachary Novack  
Deep Generative Models (CSE 291) Course Project (UC San Diego), 2022
- **Approximating Optimal Transport via GANs for Recourse Disparity Analysis**  
Zachary Novack, Qi Xuan Teo, Ryan Steed  
Probabilistic Graphic Models (10-708) Course Project (Carnegie Mellon University), 2022
- **Tracking Political Sentiment on Cold War China in Congressional Speeches**  
Zachary Novack, Eden Hu, and Mason Lin  
**1st Place** at Statistics and Data Science Research Showcase (Carnegie Mellon University), 2021
- **Lunch at the EigenSalad Bar: Linear Approaches to Dimensionality Reduction for Image Processing**  
Zachary Novack  
Numerical Linear Algebra (21-344) Course Project (Carnegie Mellon University), 2021

### *Blog Posts*

- **Armchair Statistics: Benford's Law and other Misconceptions in the Age of Data**  
Zachary Novack  
Carnegie Mellon University Triple Helix, 2021

## WORK EXPERIENCE

- AI/ML Intern* Summer 2020 - Spring 2021  
[Unisys Corporation](#), Blue Bell, PA
- Designed python implementation of categorical distance metrics to interface with scikit-learn clustering algorithms
  - Deployed time-series models (ARIMA, LSTM, Facebook Prophet) to predict computer resource utilization
  - Developed model retraining infrastructure to automatically track distribution shift in time-series models
- Front Ensemble Coordinator* Fall 2019 - Summer 2020  
[Gateway Senior High School](#), Monroeville, PA
- Led rehearsals and designed pedagogical structure for the front ensemble (non-mobile percussion) in Gateway's marching band and indoor percussion programs, working with a group of 10-15 students from ages 14-18.
- Studio Intern* Summer 2019  
[Joy Records](#), Tel Aviv, Israel
- Analyzed commercial streaming data to construct customized playlists for clients
  - Assisted in website development for Hebrew-to-English translations
- Percussion Arranger* Fall 2018 - Spring 2019  
[Tomball High School Indoor Percussion](#), Tomball, TX
- Arranged musical production for large percussion ensemble in order to compete in the Winter Guard International (WGI) national circuit

## TEACHING EXPERIENCE

- Teaching Assistant*  
 Carnegie Mellon University, Pittsburgh, PA
- *10-600: Machine Learning Primer* Summer 2022
    - Designed and demoed online course structure on prerequisite knowledge for further machine learning courses
    - Topics Covered: Basic Python programming, coding linear algebra, Big-O notation, basic calculus, IDE set-up and environment management
  - *10-301/601: Introduction to Machine Learning* Fall 2021 - Present
    - Spearheaded team maintaining autograder implementation for coding assignments
    - Led recitation and designed homework questions for class of 500+ students
    - Topics Covered: Decision Trees, Linear & Logistic Regression, Regularization, Dense and Convolutional Neural Networks, PAC Learning, Generative Models, MAP Estimation, Bayesian Networks, Hidden Markov Models, Markov Decision Processes, Clustering, Ensemble Methods
  - *85-340: Research Methods for Social Psychology* Fall 2021
    - Fully created and taught course module introducing R for psychology students, including computer science fundamentals and applications for experiment design and data analysis

- Topics Covered: Basic types, functions, vectorized programming, workflow in dplyr, basic statistical analysis, one-way and two-way ANOVA
- *36-225: Introduction to Probability Theory* *Summer 2021*
  - Topics Covered: Basic probability, random variables, univariate/multivariate probability distributions, moment-generating functions, central limit theorem
- *36-226: Introduction to Statistical Inference* *Spring 2021*
  - Topics Covered: Maximum likelihood estimation, method of moments, large & small sample hypothesis testing, properties of point estimators, confidence intervals, order statistics, Type I & Type II errors, ANOVA
- *88-300: Programming for Social Scientists* *Summer 2020 - Spring 2021*
  - Topics Covered: Basic data analysis, workflow in dplyr, basic text analysis, linear regression

## MUSIC PROJECTS

- RoboPierre* Spring 2020  
 Adaptive Impressionist Music via Generative Modeling
- Developed interactive web app to randomly generate polyphonic music trained on impressionistic composers
  - Implemented using Google Magenta’s Polyphony RNN and custom stochastic voice leading algorithm

- ThereMyn* Spring 2019  
 Motion-Controlled Monophonic Synthesizer
- Used infrared distance monitor to drive audio signal creation
  - Created front-end GUI to translate audio signals into a usable motion-controlled synthesizer

## ACCOLADES

- Honors Programs*
- *Phi Beta Kappa*, October 2021 - Present
  - *Andrew Carnegie Society Scholar*, September 2021 - Present
  - *Quantitative Social Science Scholar*, August 2018 - May 2022
  - *Dean’s List: High Honors*, December 2018 - May 2022

- Awards*
- *Small Undergraduate Research Grant (SURG)* for “Statistical Inference of On-line Radicalization in Extremist Communities”, Carnegie Mellon University, June 2021
  - *Dietrich Senior Honors Research Fellowship* for “Autoregressive Models of On-line Addiction”, Dietrich College, Carnegie Mellon University, May 2021
  - *First Place: Statistics & Data Science Research Showcase*, for “Tracking Political Sentiment on Cold War China in Congressional Speeches”, Carnegie Mellon University, May 2021
  - *Summer Undergraduate Research Fellowship (SURF)*, for “Empirical Test of the Dual Accumulator Model”, Carnegie Mellon University, June 2020
  - *Second Place: 15-112 Term Project Showcase* for “ThereMyn: Motion-Controlled Monophonic Synthesizer”, School of Computer Science, Carnegie Mellon University, April 2019

#### *Scholarships*

- *National Science Foundation Graduate Research Fellowship*, honorary mention, Spring 2022
- *Paul Mellon Memorial Presidential Scholarship* (merit-based), August 2018 - Present

## **SKILLS**

#### *Programming Languages and Packages*

- Python (Pytorch, Tensorflow, Scikit-Learn, PySpark, CVXPY), R (dplyr, tscount, zoo), C, Matlab, SQL (postgres, MySQL), Stan, Git, Shell, Max/MSP/Jitter

#### *Other Skills*

- AWS (S3, EC2, EMR), Microsoft Azure, Docker, Agile, Jira, Grafana, Ableton Live

## **EXTERNAL ACTIVITIES**

*Statistics Department Student Representative* Spring 2022 - Present  
[Dietrich College Council](#), Carnegie Mellon University, Pittsburgh, PA

- Took part in monthly council meetings to deliberate on proposals for small-to-large scale changes to Dietrich curricula and other college activities

*Professional Event Coordinator* Spring 2021 - Present  
American Statistical Association, Carnegie Mellon University, Pittsburgh, PA

- Coordinated multi-part speaker series featuring both faculty and external researchers
- Facilitated peer-mentorship program within the Statistics environment for future course planning

*Staff Writer* Fall 2020 - Present  
[The Triple Helix](#), Carnegie Mellon University, Pittsburgh, PA

- Wrote journal articles on wide-scale statistical literacy and societal impacts of misreporting experimental results

*Performer and Composer* Spring 2019 - Spring 2020  
[Exploded Ensemble](#), Carnegie Mellon University, Pittsburgh, PA

- Designed large-scale Max/MSP programs for multimedia interactive performances
- Composed electro-acoustic pieces for mixed instrumentation ensembles