

Erica Cai

ecai@cs.umass.edu

<https://ec769.github.io/>

## RESEARCH INTERESTS

---

Currently interested in natural language processing, machine learning, probabilistic graphical models, computational social science, data science, statistical inference, and causal inference research.

## EDUCATION

---

**University of Massachusetts, Amherst MA**

2020 - present

MS/PhD, Computer Science (GPA: 4.0/4.0)

**Rutgers University Honors College, New Brunswick NJ**

2016 - 2020

B.S, Computer Science major and Political Science minor (GPA: 4.0/4.0)

## RESEARCH EXPERIENCE

---

**Statistical Social Language Analysis Lab, University of Massachusetts Amherst** Fall 2022 - present

Advisor: Brendan O'Connor

Research Area: Natural Language Processing, Computational Social Science

*Developing zero-shot learning methods for event extraction among multiple agents from text in the international relations setting. In conversational text, developing methods and exploring the causal effect of gender on interruptions of advocates by Supreme Court justices in the presence of mediator variables.*

**Knowledge Discovery Lab, University of Massachusetts Amherst** Fall 2020 – Summer 2022

Advisor: David Jensen

Research Area: Causal Inference, Probabilistic Graphical Models, Explainable Artificial Intelligence

*Developing a much more efficient version of the PC algorithm, which is an algorithm that learns causal structure. Developing a method, minimum conditional dependence distribution divergence (MC3D), and using it to find that the empirical data generation process violates assumptions of algorithms that learn causal structure. Developing counterfactual explanations for reinforcement learning agent actions in an Atari game. Developing a model to predict the competency of a reinforcement learning agent.*

**University of Massachusetts Amherst**

Summer 2022

Advisor: Laura Balzer

Research Area: Causal inference, Biostatistics

*Improving on an adjustment approach for baseline covariates to maximize empirical efficiency in large randomized trials.*

**Rutgers University**

2019 – 2020

Advisor: Eric Allender

Research Area: Computational Complexity

*Exploring progress toward proving that NP is not contained in  $P_{poly}$ , a statement which would have significant implications in the cryptography field, by researching non-containment relationships between complexity classes that are slightly larger or smaller than NP and  $P_{poly}$ .*

**Center for Discrete Mathematics & Theoretical Computer Science (DIMACS)** Summer 2019

Advisor: Janne Lindqvist

*Studying public ephemeral messaging and the usage of these messages to find potential security and privacy issues, developing a scraping service to collect public messages, applying ML and heuristic algorithms to categorize messages, and using MySQL queries to analyze large dataset.*

**Bessemer Trust**

Summer 2018

Advisor: Brian Skarbek

*Designing the company's first prototype chatbot to answer employees' queries using Amazon Web Services, Python and MySQL.*

**Memorial Sloan Kettering Cancer Center Shuman Lab**

Summer 2014, 2015

Advisor: Liang Deng

*Conducting experiments to study immunotherapy efficacy in cancer treatment, particularly that of the Myxoma virus, by performing infections and titrations on mouse cells. Analyzing experimental data using software tools such as GraphPad Prism.*

**Alcatel-Lucent Bell Labs**

Fall 2015

Advisor: David Neilson

*Writing computer programs to manipulate patterns of light on an LCOS micro-display. Designing a Liquid Crystal on Silicon (LCOS) micro-display system.*

**PUBLICATIONS AND REPORTS**

---

**Improving the Efficiency of the PC Algorithm by Using Model-Based Conditional Independence Tests.** Erica Cai, Andrew McGregor, David Jensen. *NeurIPS Causal Machine Learning for Real World Impact Workshop 2022 (accepted)*

**Quantifying the Effect of Gender on Interruptions in Supreme Court Oral Arguments.** Katherine Keith, Erica Cai, Ankita Gupta, Brendan O'Connor, Douglas Rice. Under submission and accepted as abstract at Text as Data 2022.

**MC3D: A Posterior Predictive Check for Learning Directed Graphical Models.** David Jensen, Erica Cai. Under submission, 2022.

**Adaptive Selection of the Optimal Strategy to Improve Precision and Power in Randomized Trials.** Laura B. Balzer, Erica Cai, Lucas Godoy Garraza, and Pracheta Amaranath. Under submission, 2022.

**Measuring Interventional Robustness in Reinforcement Learning.** Katherine Avery\*, Jack Kenney\*, Pracheta Amaranath, Erica Cai, David Jensen. Under submission, 2022.

**Survey on Coming Closer to Finding a Relationship between NP and  $P_{poly}$ .** Erica Cai. Rutgers University Honors Scholar Report, 2020.

**A Large-Scale Study of Ephemeral Messaging Use.** Erica Cai. DIMACS Report, 2019.

## TEACHING EXPERIENCE

---

UMass COMPSCI 689: Machine Learning (Teaching Assistant)	Fall 2022
UMass COMPSCI 688: Probabilistic Graphical Models (Teaching Assistant)	Spring 2022
Rutgers COMPSCI 112: Data Structures (Recitation Instructor and Grader)	Spring 2019, 2020
Rutgers COMPSCI 206: Discrete Structures II (Recitation Instructor and Grader)	Fall 2019

## VOLUNTEERING EXPERIENCE

---

A2E Tutor at Roosevelt Elementary School, New Brunswick	2019
Rutgers University Habitat for Humanity Volunteer	2017 - 2018
Rutgers University Honors College Ambassador	2017
Brookdale Community College Summer Camp Instructor	2013 - 2014

## AWARDS

---

Edward Riseman and Allen Hanson Scholarship, UMass.	2021 - 2022
CICS Scholarship, UMass.	2020 - 2021

Inductee, Matthew Leydt Society of Rutgers Univ.	2020
Dean's Excellence Award, Rutgers Univ. SAS	2020
NSF Grant CCF-1852215, DIMACS REU Program	Summer 2019
Inductee, Phi Beta Kappa Honor Society	2019
Trustee Award Scholarship, Rutgers Univ.	2016 – 2020
3 <sup>rd</sup> Place in Northeast US and Research Grant Awarded, Young Science Achievers Program	2014, 2015
2 <sup>nd</sup> Place in Fashion Design, NJ Technology Student Association	2014

## **RELEVANT GRADUATE COURSES**

---

Machine Learning (theory-based), Probabilistic Graphical Models, Advanced Algorithms, Advanced Logic in Computer Science, Mathematical Statistics (probability theory), Natural Language Processing, Mathematical Statistics (statistical estimation), Theory of Computation seminar, Introduction to Causal Inference, Simulation and Causal Modeling seminar, Research Methodologies, Advanced Information Assurance

## **TECHNICAL SKILLS**

---

Python, R, Java, C and C++, Amazon Web Services, LaTeX, PyTorch, Scikit-Learn, Numpy, Computer-Aided Design, Spring Boot, Scheme, MySQL, Scrapy, Photoshop, Flash, HTML, Angular, NodeJS, NoSQL Databases, Microsoft Office, GraphPad Prizm, ni-max