Erica Cai
ecai@cs.umass.edu
https://ec769.github.io/

RESEARCH INTERESTS

I am very interested in and have research expertise related to knowledge graphs, natural language processing, machine learning, network science, probabilistic graphical models, computational social science, data science, statistical inference, and causal inference.

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; Honors: Phi Beta Kappa, summa cum laude)

RESEARCH AND WORK EXPERIENCE

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

Advisor: Brendan O'Connor

Research Area: Knowledge Graphs, Natural Language Processing, Network Science, Computational Social Science

Position: Graduate Research Assistant

Developing few-shot and zero-shot methods to extract information (events, relations, fine-grained named entities) from text (sentence-level and document-level), and improving evaluation of these methods. Understanding and formalizing the effect of information extraction error on analyses (e.g., transitivity, clustering coefficient) over a knowledge graph. In conversational text, developing methods and exploring the causal effect of gender on interruptions of advocates by Supreme Court justices given mediator variables.

MIT Lincoln Laboratory May 2024 – May 2025

Advisors: Benjamin Miller and Olga Simek

Position: AI/ML Research Intern (Summer 2024); Student Technical Assistant (Fall 2024, Spring 2025)

Developing a method to rank nodes in attributed graphs by importance and uniqueness of attributes, and demonstrating interesting insights of the method in applications of biomolecule structures, social networks, and supply chain networks.

MIT Lincoln Laboratory May 2023 – May 2024

Advisors: Olga Simek and Benjamin Miller

Position: AI/ML Research Intern (Summer 2023); Student Technical Assistant (Fall 2023, Spring 2024)

Developing few-shot and zero-shot methods for fine-grained named entity recognition and relation extraction to construct knowledge graphs. Exploring how relation extraction performance affects the performance of identifying important nodes in knowledge graphs.

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

Advisor: David Jensen

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

Position: Graduate Research Assistant

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 Prizm.

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.

SUBMISSIONS AND IN-PROGRESS WORK

Relation Extraction for an Entire Community Network: The AFFILKG Social Register Datasets.
Erica Cai, Sean McQuade, Kevin Young, Brendan O'Connor. *Submitted, 2025.*

UniqueRank: Identifying Important and Difficult-to-Replace Nodes in Attributed Graphs.
Erica Cai, Benjamin Miller, Olga Simek, Christopher Smith. *In progress, 2024*.

Efficient Frontier for Knowledge Graph Construction with Small Language Models.
Aishwarya Malgonde, Astha Baranwal, Nithin Govindan Kutty Raghunath, Erica Cai, David Koleczek, Jonathan Larson, Robert Ness. *Submitted, 2024.*

PUBLICATIONS AND REPORTS

Identifying and Investigating Global News Coverage of Critical Events such as Disasters and Terrorist Attacks. Erica Cai*, Xi Chen*, Reagan Keeney, Ethan Zuckerman, Brendan O'Connor, Przemyslaw Grabowicz. *International AAAI Conference on Web and Social Media 2025.*

The State of Relation Extraction Data Quality: Is Bigger Always Better? Erica Cai, Brendan O'Connor. *Association of Computational Linguistics (ACL) (Findings) 2024.*

From Low Resource Information Extraction to Identifying Influential Nodes in Knowledge Graphs.
Erica Cai, Olga Simek, Benjamin Miller, Danielle Sullivan-Pao, Evan Young, Christopher Smith. *Complex Networks XV, published in Springer Nature*, presented at International Conference on Complex Networks, 2024. (Oral)

Adaptive Selection of the Optimal Strategy to Improve Precision and Power in Randomized Trials.
Laura B. Balzer, Erica Cai, Lucas Godoy Garraza, Pracheta Amaranath. *Biometrics, 2023.*

A Monte Carlo Language Model Pipeline for Zero-Shot Sociopolitical Event Extraction. Erica Cai, Brendan O'Connor. *New England NLP (from abstract) 2023 (Oral); Text as Data (from abstract) 2023; NeurIPS Instruction Tuning and Instruction Following Workshop 2023.*

Evaluating Zero-Shot Event Structures: Recommendations for Automatic Content Extraction (ACE) Annotations. Erica Cai, Brendan O'Connor. *Association of Computational Linguistics (ACL) 2023.*

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.*

"Let Me Just Interrupt You": Estimating Gender Effects in Supreme Court Oral Arguments. Erica Cai, Ankita Gupta, Katherine Keith, Brendan O'Connor, Douglas Rice. *Journal of Law and Courts, 2024; Text as Data (from abstract) 2022.*

MC3D: A Posterior Predictive Check for Learning Directed Graphical Models. David Jensen, Erica Cai. *Submitted, 2022 and in progress.*

Measuring Interventional Robustness in Reinforcement Learning. Katherine Avery*, Jack Kenney*, Pracheta Amaranath, Erica Cai, David Jensen. *ArXiv, 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.*

INVITED PRESENTATIONS AND TALKS

Oral Presentation – From Low Resource Information Extraction to Identifying Influential Nodes in Knowledge Graphs — July 2023

International Conference on Complex Networks 2024 at Exeter, United Kingdom

Poster Presentation – A New Zero-shot Event Extraction Method — November 2023

Text as Data 2023 at UMass Amherst

Poster Presentation - Targeted Few-shot Fine-grained NER and Relation Extraction — August 2023

Graph Exploitation Symposium 2023 at MIT Endicott House

Oral Presentation – On Zero-shot Event Extraction and Uncertainty — July 2023

MIT Lincoln Lab

Oral Presentation - Efficient and Modality-Independent Zero-Shot Event Extraction of Entities with Actor Representatives — April 2023

New England NLP 2023 at UMass Lowell

TEACHING EXPERIENCE

UMass COMPSCI 685: Advanced Natural Language Processing (Teaching Assistant) — Spring 2025

UMass COMPSCI 485: Applications of Natural Language Processing (Teaching Assistant) — Fall 2023

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

SERVICE AND VOLUNTEERING EXPERIENCE

UMass NLP Webmaster 2023 - present

Industry Mentorship Course (696DS) PhD Mentor Spring 2024

ACL 2024 Reviewer / ARR Reviewer (bi-monthly) 2024

EMNLP 2023 Reviewer 2023

UMass CICS Informatics Faculty Search Committee Student Representative 2023, 2024

UMass CICS Candidate Friday Volunteer 2023

Mention by the *Strict Scrutiny* Podcast for paper on Supreme Court oral arguments 2023

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

3rd Place in Northeast US and Research Grant Awarded, Young Science Achievers Program 2014, 2015

2nd 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), Advanced Natural Language Processing, Mathematical Statistics (statistical estimation), Theory of Computation seminar, Introduction to Causal Inference, Simulation and Causal Modeling seminar, Research Methods, 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