

KATHERINE A. KEITH

kak5@williams.edu ◊ <https://kakeith.github.io/>

EDUCATION

University of Massachusetts Amherst
Ph.D. in Computer Science, September 2021
M.S. in Computer Science, February 2020

2016–2021
Amherst, MA

Lewis & Clark College
B.A. in Mathematics with departmental honors, summa cum laude
Minor: Chinese Language

2011–2015
Portland, OR

ACADEMIC APPOINTMENTS

Assistant Professor
Department of Computer Science
Williams College

July 2022–Present
Williamstown, MA

RESEARCH EXPERIENCE

Postdoctoral Young Investigator
Allen Institute for Artificial Intelligence

September 2021–June 2022
Seattle, WA

- Research with the Semantic Scholar team

Graduate Research Assistant
University of Massachusetts, Amherst

September 2016–August, 2021
Amherst, MA

- Natural language processing, machine learning, and computational social science research
- Advisor: Dr. Brendan O'Connor

Research Intern
AI Discovery Team, Bloomberg

June–August 2020
Remote (London, UK team)

- Mentors: Dr. Edgar Meij, Dr. Christoph Teichmann

Research Intern
CTO Data Science Team, Bloomberg

May–August 2018
New York, New York

- Mentor: Dr. Amanda Stent

GRANTS

- *Massachusetts Life Sciences Center's* Workforce Development Captial Grant. With Stephen N. Freund, Noah J. Sandstrom, and Victor A. Cazares. Awarded July 2024. (\$744,446 + \$83,200 matching funds)
- Young Investigator Grant from the *Allen Institute for Artificial Intelligence*. Awarded November 2022. (\$100,000)
- *Social Science Research Council (SSRC)/Summer Institutes in Computational Social Science (SICSS)* Research Grant. With Ian Stewart. Awarded 2021. (\$1,200)
- *Kaggle* Open Data Research Grant. Awarded 2020. With Andy Halterman and Sheikh Sarwar. (\$5,000)

AWARDS & HONORS

- Bloomberg Data Science Ph.D. Fellowship (2019-2021)
- Outstanding Reviewer, EMNLP 2020
- Awarded *distinction* for Ph.D. candidacy portfolio submitted to the College of Information and Computer Sciences at University of Massachusetts Amherst (December 2019)
- Empirical Methods in Natural Language Processing (EMNLP) Student Travel Scholarship (2017)
- Paul Utgoff Memorial Graduate Scholarship in Machine Learning (2016)
- Rena J. Ratte Memorial Award, Lewis & Clark College (2015)– “*This is the highest academic award given at Lewis & Clark and is given annually to one undergraduate of senior standing whose abilities and commitment have combined to produce work which is consistently of the greatest distinction.*”
- NCAA Division III Cross-Country All-Academic (2012)

JOURNAL PUBLICATIONS (PEER-REVIEWED)

2. RCT Rejection Sampling for Causal Estimation Evaluation. **Katherine A. Keith**, Sergey Feldman, David Jurgens, Jonathan Bragg, and Rohit Bhattacharya. *Transactions on Machine Learning Research (TMLR)*. 2023.
1. Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond. Amir Feder, **Katherine A. Keith**, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E. Roberts, Brandon M. Stewart, Victor Veitch, Diyi Yang. *Transactions of the Association for Computational Linguistics (TACL)*. 2022.

CONFERENCE PROCEEDINGS (PEER-REVIEWED)

9. Causal Matching with Text Embeddings: A Case Study in Estimating the Causal Effects of Peer Review Policies. Raymond Z. Zhang, Neha Nayak Kennard, Daniel Scott Smith, Daniel A. McFarland, Andrew McCallum, and **Katherine A. Keith**. In *Findings of the Association for Computational Linguistics (ACL Findings)*. 2023.
8. Words as Gatekeepers: Measuring Discipline-specific Terms and Meanings in Scholarly Publications. Li Lucy, Jesse Dodge, David Bamman, and **Katherine A. Keith**. In *Findings of the Association for Computational Linguistics (ACL Findings)*. 2023.
7. Paying Attention to the Algorithm Behind the Curtain: Bringing Transparency to YouTube’s Demonetization Algorithms. Arun Dunna, **Katherine A. Keith**, Ethan Zuckerman, Narseo Vallina-Rodriguez, Brendan O’Connor, Rishab Nithyanand. In *ACM Conference On Computer-Supported Cooperative Work And Social Computing (CSCW)*. 2022.
6. Corpus-Level Evaluation for Event QA: The IndiaPoliceEvents Corpus Covering the 2002 Gujarat Violence. Andrew Halterman*, **Katherine A. Keith***, Sheikh Muhammad Sarwar*, and Brendan O’Connor (* indicates joint first-authors). In *Findings of the Association for Computational Linguistics (ACL Findings)*. 2021.
5. Text and Causal Inference: A Review of Using Text to Remove Confounding from Causal Estimates. **Katherine A. Keith**, David Jensen, and Brendan O’Connor. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)*. 2020.
4. Modeling financial analysts’ decision making via the pragmatics and semantics of earnings calls. **Katherine A. Keith** and Amanda Stent. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL)*. 2019.

3. Uncertainty-aware generative models for inferring document class prevalence. **Katherine A. Keith** and Brendan O'Connor. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*. 2018.
2. Monte Carlo Syntax Marginals for Exploring and Using Dependency Parses. **Katherine A. Keith**, Su Lin Blodgett, and Brendan O'Connor. In *Proceedings of North American Chapter of the Association for Computational Linguistics (NAACL)*. 2018.
1. Identifying civilians killed by police with distantly supervised entity-event extraction. **Katherine A. Keith**, Abram Handler, Michael Pinkham, Cara Magliozzi, Joshua McDuffie, and Brendan O'Connor. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*. 2017.

WORKSHOP PUBLICATIONS (LIGHTLY PEER-REVIEWED)

4. Democratizing Machine Learning for Interdisciplinary Scholars: Reflections on the NLP+CSS Tutorial Series. **Katherine A. Keith** and Ian Stewart. In *Proceedings of the 1st Workshop on Teaching for NLP*. 2023.
3. Literary Intertextual Semantic Change Detection: Application and Motivation for Evaluating Models on Small Corpora. Jackson Ehrenworth and **Katherine A. Keith**. In *Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change at EMNLP*. 2023.
2. Text as Causal Mediators: Research Design for Causal Estimates of Differential Treatment of Social Groups via Language Aspects. **Katherine A. Keith**, Douglas Rice, and Brendan O'Connor. In *Proceedings of the First Workshop on Causal Inference and Natural Language Processing (CI+NLP) at EMNLP*. 2021.
1. Uncertainty over Uncertainty: Investigating the Assumptions, Annotations, and Text Measurements of Economic Policy Uncertainty. **Katherine A. Keith**, Christoph Teichmann, Brendan O'Connor, and Edgar Meij. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS) at EMNLP*. 2020.

TEACHING EXPERIENCE

Instructor

- CSCI 136: Data Structures & Advanced Programming (Williams College, Spring 2024)
2 lecture sections, 4 lab sections, 59 students, 9 TAs
- CSCI 104: Data Science and Computing For All (Williams College, Fall 2023; Co-instructor with Dr. Stephen Freund)
2 instructors splitting: 2 lecture sections, 4 lab sections, 61 students, 8 TAs
- CSCI 375: Natural Language Processing (Williams College, Spring 2023)
2 lecture sections, 46 students, 2 TAs
- CSCI 104: Understanding Data Through Computation (Williams College, Fall 2022; Co-instructor with Dr. Stephen Freund)
2 instructors splitting: 2 lecture sections, 4 lab sections, 48 students, 8 TAs
- CS335: Machine Learning (Mount Holyoke College, Spring 2020)
1 lecture section, 19 students
- CS191: Ethical Issues Surrounding Artificial Intelligence Systems and Big Data, First Year Seminar (University of Massachusetts Amherst, Fall 2019)

Graduate Teaching Assistant

- CS685: Advanced Natural Language Processing (University of Massachusetts Amherst, Spring 2018)

Online Tutorial Instructor

- *Aggregated Classification Pipelines: Propagating Probabilistic Assumptions from Start to Finish*. NLP+CSS 201 Online Tutorial Series. March 30, 2022.

Fulbright English Teaching Assistant

U.S. Department of State

August 2015–June 2016

Kinmen, Taiwan

- Taught first through sixth grade ESL courses in a public elementary school.
- Facilitated multi-cultural dialogue and events.

STUDENT RESEARCH MENTORING

Student Research Mentees

- Jacob Chen (Undergraduate Research, Summer 2023; Post-baccalaureate Research Assistant, Spring 2024)
- Justin Cheigh (Research Independent Study, Spring 2024)
- Angus Li (Undergraduate Research, Summer 2023; Research Assistant Fall 2023)
- Serah Park (Undergraduate Research, Summer 2023)
- Lucy Li (PhD intern at AI2, 2022) – *Winner of AI2's Outstanding Intern of the Year Award*
- Sirius Just (Undergraduate Honors Thesis, AY 2019-2020; REU, Summer 2019)
- Tamas Palfi (REU, Summer 2019)
- Hieu Phan (REU, Summer 2019)
- Aaron Mueller (REU, Summer 2017)

Undergraduate Honors Thesis Committee Member (Second Reader)

- Max Ellis (AY 2023-2024)
- Elijah Tamarchenko (AY 2022-2023)

SPEAKING ENGAGEMENTS

Invited Talks

- American Causal Inference Conference, “Statistical Innovations in Causal Inference with Text as Data: Emerging Trends and Future Directions” invited session. Talk entitled “Text as a High-Dimensional Causal Confounding Variable: An Empirical Framework for Evaluation” — May 16, 2024.
- Institute for Analytical Sociology (online). “Melding NLP and Causal Inference” — February 8, 2023.
- Discussant, New Directions in Analyzing Text as Data (TADA), 2023.
- University of Iowa, Computer Science Colloquium. “Melding NLP and Causal Inference” — September 8, 2023.
- Williams College Statistics Colloquium “Automated Event Extraction for News-Based Counterdata”—October 19, 2022.
- Williams College Computer Science Colloquium—November 15, 2019
- Rutgers University’s Natural Language Processing Group—November 6, 2019
- Lewis & Clark College’s Mathematical Sciences Colloquium—September 24, 2018

Guest Lectures

- Stanford University, CS 293 / EDUC 473: Empowering Educators via Language Technology. “Causal Effect Estimation with Text Data”—October 25, 2023.
- University of Massachusetts Amherst, COMPSCI 692L: Seminar, Natural Language Processing, “RCT Rejection Sampling for Causal Estimation Evaluation”—May 2, 2023
- Georgia Institute of Technology, CS-6471: Computational Social Science, “Text and Causal Estimation: Text as Causal Confounders and Mediators”—March 28, 2022
- University of Massachusetts Amherst, Natural Language Processing—November 13, 2020
- Mount Holyoke College, Natural Language Processing—November 13, 2019
- Mount Holyoke College, Natural Language Processing—October 10, 2018
- University of Massachusetts Amherst, Advanced Natural Language Processing—April 26, 2018

SERVICE & OUTREACH

Williams College Service

- Colloquium Organizer, Computer Science Department, AY 2023-2024
- Computer Science Women Organizer, Computer Science Department, AY 2023-2024
- Underrepresented Identities in Computer Science (UniCS) Faculty Mentor, Computer Science Department, AY 2022-2023

Leadership Roles in the Research Community

- Co-organizer, *NLP+CSS* workshop at NAACL 2024. June 21, 2024.
- Area Chair, *Computational Social Science and Cultural Analytics Track*, EMNLP 2023.
- Organizing Committee/Program Committee. New Directions in Analyzing Text as Data (TADA), 2023.
- Co-organizer, *NLP+CSS* workshop at EMNLP. December 7, 2022.
- Co-organizer, *NLP+CSS 201: Beyond the Basics* online tutorial series (Fall 2021)
- Co-organizer and Panel Moderator, *First Workshop on Causal Inference and Natural Language Processing* at EMNLP 2021. November 10, 2021.

Reviewer / Referee

- *NeurIPS*: 2024
- *NSF Ad-Hoc Reviewer*: 2024
- *TACL*: 2023
- *ACL Rolling Review (ARR)*: Since 2022
- *TADA*: 2022, 2023
- *Journal of Data Science*: February, 2022
- *Science Advances*: October, 2021
- *Workshop of Teaching NLP*: 2021
- *ACL*: 2020, 2021, 2023
- *NAACL*: 2019, 2021
- *EACL*: 2021
- *CoNLL*: 2020
- *EMNLP*: 2020
- *ICWSM*: 2019, 2020

Other Outreach Roles

- Panelist, Teaching at Teaching Intensive Institutions, University of Massachusetts Amherst (October 20, 2023)

- Panelist, University of Massachusetts CSWomen panel on “Applying to Research Internships and Jobs” (November 19, 2021)
- Panelist, NSF CSGrad4US Mentoring Program panel on “What I wish I knew when I started graduate school” (November 24, 2021)
- PhD Applicant Support Mentor for Underrepresented Students (November 2020) University of Massachusetts Amherst
- Panelist, CS Graduate School Q&A Panel (May 1, 2020) Cape Cod Community College
- Co-chair, CSWomen (February 2019–January 2020) University of Massachusetts Amherst Computer Science Women’s Group
- Social Co-Chair, CSWomen (January 2017–January 2019) University of Massachusetts Amherst Computer Science Women’s Group
- Student Volunteer, Women in Engineering and Computing Career Day (October 28, 2019) University of Massachusetts Amherst
- “What is computer science?” 7th Grade Guest Lecture (Jan. 3, 2018 and Jan. 7, 2019) Chief Joseph Middle School, Bozeman, Montana
- Organizer, CICS Graduate Male Ally Workshop Series (September 2017– May 2018) University of Massachusetts Amherst
- Student Volunteer, Girls Inc. Eureka! Summer Workshop (August 1 & 3, 2017) University of Massachusetts Amherst
- Student Volunteer, Women in Engineering and Computing Career Day (October 24, 2016) University of Massachusetts Amherst

PROFESSIONAL DEVELOPMENT

Workshops on Academia: Research, Teaching, and/or Mentoring

- Faculty Success Program (12-weeks online; Fall 2023) by National Center for Faculty Development and Diversity (NCFDD).—*Intensive online program over the course of twelve weeks consisting of weekly group calls, weekly homeworks, and one-on-one coaching.*
- Participated in *Claiming Williams Day* “Creating an Inclusive Classroom” Workshop with Dr. Anika Daniels-Osaze. February 2, 2023.
- Participated in the Computing Research Association (CRA)’s Career Mentoring Workshop. Washington DC, February 24-25, 2022.
- Computing Research Association of Women (CRA-W) Graduate Cohort (2017, 2018)

Roundtables

- Williams Roundtable Program: Research Discussion Group, Spring 2023. (\$500 stipend)

FOREIGN LANGUAGES

Chinese (Mandarin)

HSK Level 4 (test passed April 16, 2016)

CET Beijing: 16-week language-intensive immersion program (Spring 2014)