

KATHERINE A. KEITH

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

EDUCATION

| | |
|---|---------------------|
| University of Massachusetts Amherst | 2016–2021 |
| Ph.D. in Computer Science, September 2021 | <i>Amherst, MA</i> |
| M.S. in Computer Science, February 2020 | |
| Lewis & Clark College | 2011–2015 |
| B.A. in Mathematics with departmental honors, summa cum laude | <i>Portland, OR</i> |
| Minor: Chinese Language | |

ACADEMIC APPOINTMENTS

| | |
|--------------------------------|-------------------------|
| Assistant Professor | July 2022–Present |
| Department of Computer Science | <i>Williamstown, MA</i> |
| Williams College | |

RESEARCH EXPERIENCE

| | |
|--|--------------------------|
| Postdoctoral Young Investigator | September 2021–June 2022 |
| <i>Allen Institute for Artificial Intelligence</i> | <i>Seattle, WA</i> |

- Research with the Semantic Scholar team

| | |
|---|-----------------------------|
| Graduate Research Assistant | September 2016–August, 2021 |
| <i>University of Massachusetts, Amherst</i> | <i>Amherst, MA</i> |

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

| | |
|-------------------------------------|---------------------------------|
| Research Intern | June–August 2020 |
| <i>AI Discovery Team, Bloomberg</i> | <i>Remote (London, UK team)</i> |

- Mentors: Dr. Edgar Meij, Dr. Christoph Teichmann
- Resulted in publication at the *Workshop on Natural Language Processing and Computational Social Science (NLP+CSS) at EMNLP, 2020*.

| | |
|---|---------------------------|
| Research Intern | May–August 2018 |
| <i>CTO Data Science Team, Bloomberg</i> | <i>New York, New York</i> |

- Mentor: Dr. Amanda Stent
- Resulted in publication at the *Association of Computational Linguistics (ACL), 2019*.

JOURNAL PUBLICATIONS (PEER-REVIEWED)

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)

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.

GRANTS

- Young Investigator Grant from the Allen Institute of Artificial Intelligence (\$100,000). November 2022.
- Social Science Research Council (SSRC)/Summer Institutes in Computational Social Science (SICSS) Research Grant 2021 (\$1,200). With Ian Stewart.

- Kaggle Open Data Research Grant 2020 (\$5,000). With Andy Halterman and Sheikh Sarwar.

TEACHING EXPERIENCE

Instructor

- CSCI 375: Natural Language Processing (Williams College, Spring 2023)
- CSCI 104: Understanding Data Through Computation (Williams College, Fall 2022)
- CS335: Machine Learning (Mount Holyoke College, Spring 2020)
- 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.

RESEARCH STUDENTS & MENTEES

- Jacob Chen (Undergraduate Reserach, Summer 2023)
- Serah Park (Undergraduate Reserach, Summer 2023)
- Angus Li (Undergraduate Reserach, Summer 2023)
- Elijah Tamarchenko (Second Reader on Undergraduate Honors Thesis, AY 2022-2023)
- 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)

PROFESSIONAL DEVELOPMENT

Workshops on Teaching and Mentoring

- 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.

Roundtables

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

SPEAKING ENGAGEMENTS

Invited Talks

- University of Massachusetts Amherst, COMPSCI 692L: Seminar, Natural Language Processing, “RCT Rejection Sampling for Causal Estimation Evaluation” — May 2, 2023
- Williams College Statistics Colloquium “Automated Event Extraction for News-Based Counterdata”—October 19, 2022.
- Georgia Institute of Technology, CS-6471: Computational Social Science, “Text and Causal Estimation: Text as Causal Confounders and Mediators”—March 28, 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

- 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

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

Research Community Leadership Roles

- Co-organizer, *NLP+CSS* workshop at EMNLP. December 7, 2022.
- Co-organizer, *CS+NLP 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

Reviewer / Referee

- *TADA*: 2022
- *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, 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

GRADUATE AWARDS & HONORS

- 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)
- Bloomberg Data Science Ph.D. Fellowship (awarded May 2019)
- Computing Research Association of Women (CRA-W) Graduate Cohort (2017, 2018)
- Empirical Methods in Natural Language Processing (EMNLP) Student Travel Scholarship (2017)
- Paul Utgoff Memorial Graduate Scholarship in Machine Learning (2016)

UNDERGRADUATE AWARDS & HONORS

- Fulbright ETA Grantee with the U.S. Department of State (2015–16)
- Rhodes Scholarship Finalist (2015)
- Marshall Scholarship Finalist (2015)
- 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)
- Lewis & Clark College Cross-Country Four-Year Varsity Letter Winner (2011–2015)

FOREIGN LANGUAGES

Chinese (Mandarin)

HSK Level 4 (test passed April 16, 2016)

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