

Alexandra Schofield

Associate Professor of Computer Science
xanda@cs.hmc.edu
www.cs.hmc.edu/~xanda

Office: McGregor Hall 330
Harvey Mudd College
301 Platt Boulevard
Claremont, CA, USA 91711
Office Phone: (909) 607-2105

Education

Ph.D., Computer Science - Cornell University, Ithaca, NY May 2019
Dissertation: Text Processing for the Effective Application of Latent Dirichlet Allocation
Chair: David Mimno

B.S., Mathematics and Computer Science - Harvey Mudd College, Claremont, CA May 2013
Honors in Computer Science; honors in Humanities, Social Science and the Arts

Research and Teaching Appointments

Associate Professor - Harvey Mudd College Summer 2025-present

Assistant Professor - Harvey Mudd College Summer 2019-Summer 2025

Lecturer - Cornell University Fall 2018

External Roles

Instructor, *Text Analysis Pedagogy Institute* Summer 2022
Taught modules on text processing using Python to a mixed audience of librarians and humanists interested in text analysis.

Virtual Lecturer, *Computer Science I (Outlier.org)* Spring 2022
Recorded half the lectures for a twelve-week virtual introduction to CS course in Java, recorded as an interactive module.

Research Intern, *Microsoft Research FATE Group, NYC* Summer 2018
Developing techniques and experiments for training and recovering private latent variable models on Poisson-distributed count data, such as text documents and events.

Research Intern, *Microsoft Research Machine Teaching Group, Redmond* Summer 2016
Worked with software engineers to design, implement, and user test a workflow for users to interactively debug text-based machine learning models using dictionary features.

Courses Taught or Assisted

†Co-taught with another instructor

Harvey Mudd College

Instructor, **CS 5: Introduction to Computer Science** Fall 2021, Spring 2022, Fall 2026

Instructor, **CS 123: Computing Practice, Projects, and People** Spring 2023, Fall 2023, Spring 2024, Fall 2024, Spring 2025

Instructor, **CS 140: Algorithms** Fall 2019, Spring 2020, Fall 2020, Fall 2022, Fall 2024, Spring 2025

Instructor, **CS 159: Natural Language Processing** Fall 2020, Spring 2021, Fall 2021, Spring 2024

Instructor, **CS 183: Computer Science Clinic** Fall 2019-Fall 2024

Guest lectures: HUM 340 CGU: Digital Humanities (Fall 2019) · MATH 189 HMC: Mathematical Data Science and Topic Modeling (Spring 2022) · CS 158 HMC: Machine Learning (Spring 2024)

Cornell University

Instructor, CS 4820: Analysis of Algorithms	Fall 2018 [†]
Guest Lecturer, CS 4820: Analysis of Algorithms	Fall 2017
Guest Lecturer, CS 4740: Natural Language Processing	Fall 2015, Fall 2016
Teaching Assistant, INFO 4500: Language and Technology	Spring 2016
Instructor, INFO 3350: Text Mining for History and Literature	Fall 2015 [†]

Selected Academic Honors

NDSEG Fellowship	2016
NSF GRFP Fellowship (declined)	2016
Anita Borg Memorial Scholarship	2016
MSR Graduate Women's Scholarship	2015
CRA Outstanding Undergraduate Research Award Finalist	2013

Publications and Presented Work

*represents mentored students

Archival Publications

S. Rivoire and **A. Schofield**. "WIP: Revealing the Hidden Curriculum: A Multi-Institutional Perspective." Accepted for presentation at *Frontiers in Education*, November 2025.

A. Schofield, S. Wu*, T. Bayard de Volo*, T. Kuze*, A. Gomez*, S. Sultana. " 'My Very Subjective Human Interpretation': Domain Expert Perspectives on Navigating the Text Analysis Loop for Topic Models." *Proceedings of the ACM on Human-Computer Interaction* (presented at GROUP '25), Volume 9, Issue 1. January 2025.

N. Phillips*, S. Anand*, M. Lum*, Manisha Goel, Michelle Zemel, **A. Schofield**. "Cheap Talk: Topic Analysis of CSR Themes on Corporate Twitter." *Proceedings of the Joint FinNLP-KDF-ECONLP Workshop at LREC-COLING*, May 2024.

A. Schofield. "The Possibilities and Limitations of Natural Language Processing for the Humanities." Chapter in *The Bloomsbury Handbook to the Digital Humanities*, November 2022.

J. Cheevaprawatdomrong*, **A. Schofield**, Attapol Rutherford. "More Than Words: Collocation Retokenization for Latent Dirichlet Allocation Models." *Findings of the ACL*, May 2022.

M. Goel, **A. Schofield**, M. Zemel. "One Shock, Many Disruptions: Firm Experience After India's Demonetization." *AAAI Workshop on Knowledge Discovery from Unstructured Data in Financial Services*, March 2022.

J. Hessel and **A. Schofield**. "How Effective Is BERT Without Word Ordering? Implications for Language Understanding and Data Privacy." *ACL-IJCNLP*, August 2021.

A. Schofield, R. Wicentowski, J. Medero. "Learning How To Learn NLP: Developing Introductory Concepts Through Scaffolded Discovery." *NAACL Workshop on Teaching NLP*, June 2021.

A. Eads, **A. Schofield**, F. Mahootian*, D. Mimno, R. Wilderom. "Separating the Wheat from the Chaff: A Topic- and Keyword-Based Procedure for Identifying Research-Relevant Text." *Poetics*, June 2021.

A. Schein, Z. S. Wu, **A. Schofield**, M. Zhou, H. Wallach. "Locally Private Bayesian Inference for Count Models." *ICML*, June 2019.

A. Schofield, L. Thompson, D. Mimno. "Quantifying the Effects of Text Duplication on Semantic Models." *EMNLP*, September 2017.

A. Schofield, M. Magnusson, D. Mimno. "Pulling Out the Stops: Rethinking Stopword Removal for Topic Models." *EACL*, April 2017.

A. Schofield, D. Mimno. "Comparing Apples to Apple: The Effects of Stemmers on Topic Models." *TACL* Vol. 4, July 2016. Presented at *EMNLP*, Nov 2016.

R. Keller, **A. Schofield**, A. Toman-Yih, Z. Merritt, & J. Elliot. "Automating the Explanation of Jazz Chord Progressions using Idiomatic Analysis." *MIT Computer Music Journal*, Winter 2013.

R. Keller, Z. Merritt, **A. Schofield**, & A. Toman-Yih. "A Creative Improvisation Companion Based on Idiomatic Harmonic Bricks." *Proceedings of the ICCM*, Dublin, Ireland, May 2012.

Non-Archival Peer-Reviewed Works

- S. Fulcar*, C. Ashmun*, J. Baaken*, A. Ding*, J. P. Walker*, **A. Schofield**, S. Kavassalis. "Knowing Best: Distinguishing Features of Scientific Papers Cited in Clean Air Policy." Poster at the *Annual Meeting of the American Geophysical Union (AGU)*, December 2023.
- N. Phillips*, M. Lum*, Manisha Goel, Michelle Zemel, **A. Schofield**. "Cheep Talk: Investigating Corporate Social Responsibility Speech on Twitter." Poster at *New Directions in Analyzing Text as Data*, November 2023.
- S. Babb*, M. Celeste*, D. Harris*, I. Wu*, T. B. de Volo*, A. Gomez*, T. Kuze*, T. Lee*, D. Mimno, **A. Schofield**. "Introducing tSLDA: A Workflow-Oriented Topic Modeling Tool." Poster at *West Coast NLP Workshop*, October 2021.
- C. Mallinson, M. Cahill, E. Francis, S. Namboodiripad, J. E. R. Labrada, E. M. Bender, **A. Schofield**. "Teaching Ethics across the Linguistics Curriculum." Accepted panel at *LSA Annual Meeting*, January 2021.
- T. B. de Volo*, A. Gomez*, T. Kuze*, **A. Schofield**. "LDA in the Wild: How Practitioners Develop Topic Models." Poster at *West Coast NLP Workshop*, October 2020.
- E. M. Bender, D. Hovy, **A. Schofield**. "Integrating Ethics in the NLP Classroom." Tutorial at *ACL 2020*, July 2020.
- R. Bommasani, S. Wu, **A. Schofield**. "Toward Synthetic Private Text Generation." Poster at *NeurIPS Workshop on Machine Learning with Guarantees*, December 2019.
- A. Schofield**, G. Yauney, D. Mimno. "Combatting The Challenges of Local Privacy for Distributional Semantics with Compression." Accepted Poster at *NeurIPS Workshop on Privacy in Machine Learning (PriML)*, December 2019.
- A. Schofield**, M. Magnusson, L. Thompson, D. Mimno. "Understanding Text Pre-Processing for Latent Dirichlet Allocation." Poster at *ACL Workshop for Women in NLP (WiNLP)*, June 2017.
- A. Schofield**, L. Mehr. "Gender-Distinguishing Features in Film Dialogue." Talk at *NAACL 2016 Workshop on Computational Linguistics for Literature (CLfL)*, June 2016.
- J. Hessel, **A. Schofield**, L. Lee, & D. Mimno. "What do Vegans do in their Spare Time? Latent Interest Detection in Multi-Community Networks." Poster at *NeurIPS Workshop on Networks*, December 2015.

Grants and Funding

Team Member and Workshop Coordinator, Harvey Mudd College Innovation Accelerator Laboratory for Data Science for Social Impact.

Ran 2025-2026, a local initiative focused on advancing data science practices for real world, socially-motivated problems. Total allocated: \$240,000.

Principal Investigator, NSF #2243941: REU Site: Exploring the Limits of Intelligent Systems.

Ran 2023-2026, funding 10 students per summer over 3 summers. Total allocated: \$400,097. Co-PI: George Montañez.

Presentations

"Why Isn't Computational Text Analysis a Piece of Cake?" Language Technologies Institute Colloquium at Carnegie Mellon University, October 2022.

"When Text Mining Isn't a Piece of Cake." Cognitive Science Seminar at University of Memphis (Virtual), September 2021.

"Text Analysis Isn't a Piece of Cake." Stauffer Lecture at Harvey Mudd College (Virtual), June 2021.

"Topic Models Aren't a Piece of Cake." CS/ECE Seminar at U. Mass. Dartmouth (Virtual), October 2020.

"When Text Mining Isn't a Piece of Cake." Invited talk at Women in Machine Learning (WiML) 2019.

"Toward Practical and Locally Private Inference of Topics." Invited talk at UC Irvine's Machine Learning Seminar Series, November 2019.

"How I Made This: Topic Models in the Classroom." Invited talk for Digital Humanities at the Claremont Colleges Summer Institute. August 2019.

"Topic Models and Real World Text." Invited talk at Cornell Library. May 2019.

"7 Lessons from Teaching (That I wish I'd Learned Earlier)". Invited talk for Cornell Computer Science Brown Bag Speaker Series. February 2019.

A. Schofield, "Toward Practical and Locally Private Inference of Topics." *Conference on New Directions in Analyzing Text as Data (TADA)*, September 2018.

Miscellaneous

S. Anand*, Q. Daedal, **A. Schofield**. "DSC #20: Xanda Rescues the Topic Modeling Disaster." *The Data-Sitter's Club*, November 2023. Online at <https://datasittersclub.github.io/site/dsc20.html>.

A. Schofield, T. Davidson. "Identifying Hate Speech in Social Media." Published tutorial. *ACM Crossroads Magazine (XRDS)* Vol 24 (2), December 2017. <http://www.cs.cornell.edu/~xanda/interviews.html>.

A. Schofield. "Compiler Code Comprehension: Using Inference to Supplement Prior Knowledge." *Consortium of Computer Science Colleges: Southwest (CCSC:SW)*, April 2013.

Undergraduate Student Mentorship

WHISK Lab: Workflows for Humanistic Inference of Statistical Knowledge

Simon Babb (Haverford '22), Theo Bayard de Volo (Pitzer '22), Lewis Bowen (HMC '25), Mia Celeste (HMC '24), Alfredo Gomez (HMC '21), Dana Harris (CMC '22), Stephanie Huang (HMC '24), Tatsuki Kuze (HMC '22), Taeyun Lee (HMC '21), Jean Pierre Nizyumukiza (HMC '23), Nanako Noda (HMC '23), Jamie Tang (Williams '25), Nevaeh Thompson (HMC '25), Heather Wei (Berkeley '25), Ingrid Wu (HMC '23), Qing Yang (HMC '23)

EconText: Text Analysis for Economics Applications

Deen Amanat (Pomona '24), Sathvika Anand (HMC '24), Eloise Burtis (Pomona '24), Nathaniel Getachew (Pomona '23), Samika Goel (Pomona '24), Matthew Ivler (Pomona '22), Michelle Lum (HMC '23), Julien Luebbers (Pomona '23), Sam Malik (Pomona '24), Christy Marchese (Pomona '24), Chris Nardi (Pomona '22), Nile Phillips (HMC '24)
Co-PIs: Manisha Goel, Michelle Zemel

Independent Study Projects Supervised

Policy Documents in Climate: Ford Ashmun (HMC '24), Jeremy Bakken (HMC '23), Anna Ding (HMC '23), Stephanie Fulcar (HMC '25), JP Walker (Scripps '25) (with PI Sarah Kavassalis)

Environmentalism in Psychology: Ingrid Tsang (HMC '21) (with PI Adam Pearson)

Scots Wikipedia: Emily Chen (HMC '22), Serenity Wade (HMC '21)

Gender in Language: Ben Hinthorne (HMC '21)

Privacy-Preserving N-Gram Generation: Rishi Bommasani (Cornell '19)

Independent Text Corpus Explorations: Terence Chen (HMC '26), Aidan Deshong (HMC '28), Elsa Li (HMC '28), Siri Xu (HMC '27), Theo Bayard de Volo (Pitzer '22), Nisha Bhatia (HMC '21), Violet Burbank (HMC '21), Alfredo Gomez (HMC '21), Teresa Ibarra (HMC '20), Taeyun Lee (HMC '21), Charlotte LeMay (HMC '21), Rui-Jie Yew (Scripps '21), Ashrita Raman (Cornell '21), Fauna Mahootian (Cornell '19), James Chen (Cornell '17)

Service

Leadership and Coordination

Faculty Executive Committee Representative, Harvey Mudd College	AY 2024-2025
CS Dept. Summer Research Coordinator, HMC Computer Science	Summer 2023, 2024
Affiliate Co-coordinator, Aspirations in Computing for NCWIT Inland Empire Chapter	Fall 2021-Summer 2024
Co-organizer, Widening NLP Workshop	Winter 2020-Winter 2022
Advisor, ACM-W Claremont Colleges Chapter	Fall 2019-present
Co-chair, Ethics Advisory Committee at ACL 2021	Spring-Summer 2021

Peer Review

Conferences

Association of the Advancement of Artificial Intelligence (AAAI)

Association of Computational Linguistics: Annual Meeting (ACL), European Chapter (EACL), North American

Chapter (NAACL), Rolling Review (ARR)

Conference on Computer-Supported Cooperative Work & Social Computing (CSCW)

Empirical Methods in Natural Language Processing (EMNLP)

Conference on Supporting Group Work (GROUP)
 International Conference on Machine Learning (ICML)
 Neural Information Processing Systems (NeurIPS)
Ethics Reviewer for: COLM, *ACL/ARR, NeurIPS

Black in AI (BAI)
 New in Machine Learning (co-located with NeurIPS)
 TeachNLP
 Widening NLP (WiNLP)
 Women in Machine Learning (WiML)

Workshops

Abusive Language Workshop (ALW)
 Adv. in Approx. Bayesian Inference (AABI)
 Black in AI (BAI)
 New in Machine Learning (co-located with NeurIPS)
 TeachNLP
 Widening NLP (WiNLP)
 Women in Machine Learning (WiML)

Journals and Publishers

Manning Publications
 Nature Machine Intelligence
 PeerJ Computer Science
 Science Advances

Workshops

Abusive Language Workshop (ALW)
 Adv. in Approx. Bayesian Inference (AABI)

Awards and Grants

CRA Outstanding Undergraduate Research Awards
 National Endowment for the Humanities

Outreach Workshops

<i>Workshop Leader, SWE WEST</i>	Spring 2024
<i>Taught a 75-minute session on the history of text encodings for high school girls</i>	
<i>Workshop Leader, Expanding Your Horizons</i>	Spring 2015, Spring 2016, Spring 2017
<i>Taught 1.5 hour 1-session workshops for 7th-9th grade girls on state machines, network protocols, and computer animation</i>	
<i>Teaching Assistant, CURIE Academy</i>	Summer 2015
<i>Led exercises and supported labs for 1 week of instruction in computer graphics for high school girls</i>	