

David A. Jurgens

CONTACT INFORMATION	UMSI Office 3385 Ann Arbor, MI 48109	<i>E-mail:</i> jurgens@umich.edu <i>WWW:</i> http://jurgens.people.si.umich.edu/
EDUCATION	University of California, Los Angeles , Ph.D. in Computer Science, June 2014 Washington University in St. Louis , M.S. in Computer Science, December 2004 and B.A. in Philosophy and Political Science, May 2004	
ACADEMIC EMPLOYMENT	University of Michigan School of Information Assistant Professor, August 2017 – August 2023 Associate Professor, August 2023 – <i>present</i> Computer Science and Engineering Associate Professor (joint-appointment), August 2023 – <i>present</i> Stanford University, NLP and SNAP Groups Postdoctoral Scholar, August 2015 – August 2017 McGill University, Network Dynamics Lab Postdoctoral Scholar, June 2014 – July 2015 Sapienza University of Rome, Linguistics Computing Laboratory Research Scientist, Jan. 2013 – Feb. 2014	
FUNDING AND GIFTS	Department of Education “Methods and Software to Classify College Courses At Scale” \$899,226. Co-PI with Kevin Stange (PI) and Allyson Flaster (co-PI). (09/01-2024– <i>present</i>). University of Michigan “Midwest Speech & Language Days” ADVANCE Elizabeth Caroline Crosby Faculty Grants Program (\$12,740). PI along with Co-PIs Joyce Chai, Rada Mihalcea, Emily Mower-Provost, and Justine Zhang. 04/2024 University of Michigan “Addressing Technical and Ethical Issues in Information Retrieval” Sole PI. Gilbert Whitaker Fund for the Improvement of Teaching. (\$5,942) 05/2023-03/2024 LG AI Research “Towards People-Centric Language Models with Self Alignment and Prompt Refinement” \$163,636. PI with Co-PI Rada Mihalcea. (11/2023– <i>present</i>) NIH “Mixed methods examination of warning signs within 24 hours of suicide attempt in hospitalized adults” \$3,117,711. Co-I with PI Courtney Bagge, Co-PI Kenneth Conner, Co-I Cheryl King, Co-I Paul Pfeiffer, Co-I Anthony Pisani, and Co-I Andrew Littlefield. (2023– <i>present</i>) Prolific.co “Demographically-aware Crowdsourcing.” \$10,000. Sole PI. (04/2023– <i>present</i>) DSO National Laboratories “Understanding Cultural Similarity through Mapping Moral Identity Across Nations.” \$352,721. Sole PI. (03/2023– <i>present</i>) IARPA “Stylometric Authorship Discernment and Interpretation for Realistic Inputs (SADIRI)” \$1,096,437 (U-M portion). Sole PI at U-M, with ISI prime Elizabeth Boschee (PI) and Scott Miller (co-PI) and other PIs at the University of Maryland and University of Birmingham (UK). (9/2022– <i>present</i>) LG AI Research “Knowledge-grounded Personalized Dialog Generation” \$200,000. PI with Co-PI Rada Mihalcea. (9/2022– <i>present</i>)	

NSF “CAREER: Fostering Prosocial Behavior and Well-Being in Online Communities” \$597,983. Sole PI. (10/2022–*present*)

Sandia National Labs “Developing a relationship aware disinformation flow model for testing data analytic methods” \$52,000. Co-PI with PI Daniel Romero. (03/2022–09/2022)

NIH “Aging, Major Life Transitions, and Suicide Risk” \$3,883,148. Co-I with PI Briana Mezuk, MPI Kara Washington, Co-I Sarah Burgard, Co-I Jacqui Smith, Co-I Philippa Clark, Co-I Jessica Faul, Co-I Amanda Sonnega, and Co-I Courtney Bagge. (2021–*present*)

NSF “NSF Convergence Accelerator Track F: Misinformation Judgments with Public Legitimacy” \$750,000. Co-PI with PI Paul Resnick and Co-PI Amy X. Zhang (U. Washington). (2021–2022)

University of Michigan “Measuring Racial Disparity in the Language of Physician-Patient Interactions” PI with co-PI Allison Earl (U-M Psychology) \$29,628 over one year (08/2021–*present*)

American Foundation for Suicide Prevention “Suicide Risk During Employment and Housing Transitions in Mid- and Late-Life” Co-PI with PI Briana Mezuk (U-M School of Public Health) and co-PI Jacqui Smith (U-M Psychology) \$124,880 (08/2020–7/2022)

NSF “CHS: Small: The Ties that Relate Us: Modeling the Impact of Relationships on Social Contagion and Network Dynamics” \$499,925 over three years. PI with co-PI Daniel Romero. (2020–*present*)

NSF “CRII: RI: Explainable Recognition of Social Relationships from People’s Linguistic Interactions” \$174,999 over two years (2019–03/2022); Sole PI; REU Supplement \$24,000

Amazon “Amazon Alexa Prize Finalist: Team Audrey” \$250,000 over one year (09/2019–05/2020)

NVIDIA Hardware grant for 1 Titan V GPU for deep learning. 2019.

University of Michigan “MCubed: Measuring the Impact of Early Stage Characteristics on Long Term Success of Online Communities” \$60,000 (05/2019–06/2020)

Volkswagen Foundation Co-Principal Investigator, with Fabian Flöck (GESIS), Przemyslaw Grabowicz (Max Planck), and Scott Hale (Oxford). “Current Affairs 2.0: Agenda Setting in the European Union.” Call for Computational Social Science. 842,900 € over four years. (2017–2022)

FigureEight Co-Principal Investigator, with Dan Jurafsky (Stanford) and Yulia Tsvetkov (CMU). “Fostering Civility in Online Communication.” \$175,000, one year. (2018)

AWARDS Honorable Mention at the International Conference on Computational Social Science (IC2S2) for Wright et al. 2023 (top 5 submissions)

Best Paper at the Workshop on Social Influence in Conversations (SICoN) for Ananthasubramaniam et al. (2023)

Outstanding Methodology Award at ICWSM for Mendelsohn et al. (2023)

[Cultural Competence in Computing \(3C\) Fellow 2021](#)

Best Paper at the Workshop for Noisy User-generated Text (W-NUT) for Ghosh et al. (2021)

Best Paper Runner Up at Web Science for Im et al. (2020)

Best Paper Nomination at SocInfo for Dong, Banea, Jurgens, and Mihalcea (2019)

Best Paper Award at ICWSM for Rajadesingan, Mahalingam, and Jurgens (2019)

Best Poster Presentation award at WWW for Wang et al. (2019)

Cialdini Prize from the Society for Personality and Social Psychology for our PNAS paper on the Language from Police Body Cameras (2018)

Cozzarelli Prize in Behavioral and Social Sciences from the National Academy of Sciences for our PNAS paper on the Language from Police Body Cameras (2017)

Best paper nomination at ACL for Pilehvar, Jurgens and Navigli (2013).

PUBLICATIONS ¹

Peer-Reviewed Journal Articles

Diya Yang, Dirk Hovy, **David Jurgens**, and Barbara Plank. “The Call for Socially Aware Language Technologies” *Computational Linguistics* forthcoming.

Minxue Niu, Emily Mower Provost, **David Jurgens**, Susan A Gelman, Ethan Kross, and Ariana Orvell. “The persuasive role of generic-you in online interactions” *Scientific Reports* 15 (1), 1347. 2025

Minje Choi, Daniel Romero, and **David Jurgens**. “Profile Update: The Effects of Identity Disclosure on Network Connections and Language” *EPJ Data Science*, 13(1), 45. 2024.

Aparna Ananthasubramaniam, **David Jurgens**, and Daniel Romero. “Networks and Identity Drive Geographic Properties of the Diffusion of Linguistic Innovation.” *npj Complexity*. 2024.

Hao Peng, Misha Teplitskiy, and **David Jurgens**. “Author Mentions in Science News Reveal Wide-Spread Ethnic Bias.” *Journal of Quantitative Social Science* (QSS). 2024.

Aparna Ananthasubramaniam, **David Jurgens**, Eskira Kahsay, Briana Mezuk. “Characteristics of and variation in suicide mortality related to retirement during the Great Recession: perspectives from the National Violent Death Reporting System.” *The Gerontologist*, gnae015. 2024.

Minje Choi, **David Jurgens**, and Daniel Romero. “Analyzing the Engagement of Social Relationships During Life Event Shocks in Social Media.” Proceedings of the International Conference on Web and Social Media (ICWSM). 2023.

Julia Mendelsohn, Sayan Ghosh, **David Jurgens**, and Ceren Budak. “Bridging Nations: Quantifying the Role of Multilinguals in Communication on Social Media.” Proceedings of the International Conference on Web and Social Media (ICWSM). 2023. **Outstanding Methodology Award**

Aarohi Srivastava *et al.* (100+ authors; I’m in the middle). “Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models.” *Transactions on Machine Learning Research*. 2023. [link]

Mezuk, Briana, Linh Dang, **David Jurgens**, and Jacqui Smith. “Work expectations, depressive symptoms, and passive suicidal ideation among older adults: Evidence from the Health and Retirement Study.” *The Gerontologist*. 2022.

Bas Hofstra, Daniel A. McFarland, Sanne McFarland Smith, and **David Jurgens**. “Diversifying the Professoriate.” *Socius*. 2022.

Bradley P. Turnwald, Margaret A. Perry, **David Jurgens**, Vinodkumar Prabhakaran, Dan Jurafsky, Hazel R. Markus, Alia J. Crum. “Language in Popular American Culture Constructs the Meaning of Healthy and Unhealthy Eating: Narratives of Craveability, Excitement, and Social Connection in Movies, Television, Social Media, Recipes, and Food Reviews.” *Appetite*. 2022.

Tomohiro M. Ko, Viktoryia A. Kalesnikava, **David Jurgens**, and Briana Mezuk. “A Data Science Approach to Estimating the Frequency of Driving Cessation Associated Suicide in the US: Evidence From the National Violent Death Reporting System.” *Frontiers in Public Health*. 2021.

¹Papers authored by student collaborators where I was acting in an advisory role are denoted with an underline. Papers prior to 2017 were when I was publishing as a student or postdoc.

Minje Choi, Ceren Budak, Daniel Romero, and **David Jurgens**. “More than meets the tie: Examining the Role of Interpersonal Relationships in Social Networks.” International Conference on Web and Social Media (ICWSM), 2021.

Matthew J. Salganik, Ian Lundberg, Alexander T. Kindel, Caitlin E. Ahearn, Khaled Al-Ghoneim, Abdullah Almaatouq, Drew M. Altschul, Jennie E. Brand, Nicole Bohme Carnegie, Ryan James Compton, Debanjan Datta, Thomas Davidson, Anna Filippova, Connor Gilroy, Brian J. Goode, Eaman Jahani, Ridhi Kashyap, Antje Kirchner, Stephen McKay, Allison C. Morgan, Alex “Sandy” Pentland, Kivan Polimis, Louis Raes, Daniel E. Rigobon, Claudia V. Roberts, Diana M. Stanescu, Yoshihiko Suhara, Adaner Usmani, Erik H. Wang, Muna Adem, Abdulla Alhajri, Bedoor AlShebli, Redwane Amin, Ryan B. Amos, Lisa P. Argyle, Livia Baer-Bositis, Moritz Büchi, Bo-Ryehn Chung, William Eggert, Gregory Faletto, Zhilin Fan, Jeremy Freese, Tejomay Gadgil, Josh Gagné, Yue Gaobj, Andrew Halpern-Manners, Sonia P. Hashim, Sonia A. Hausen, Guanhua He, Kimberly Higuera, Bernie Hogan, Ilana M. Horwitz, Lisa M. Hummel, Naman Jain, Kun Jin, **David Jurgens**, Patrick C. Kaminski, Areg Karapetyan, E. H. Kim, Ben Leizman, Naijia Liu, Malte Möser, Andrew E. Mack, Mayank Mahajan, Noah Mandell, Helge-Johannes Marahrens, Diana Mercado-Garcia, Viola Mocz, Katariina Mueller-Gastell, Ahmed Musse, Qiankun Niu, William P. Nowak, Hamidreza Omidvar, Andrew Or, Karen Ouyang, Katy M. Pinto, Ethan Porter, Kristin E. Porter, Crystal Qian, Tamkinat Rauf, Anahit Sargsyan, Thomas Schaffner, Landon Schnabel, Bryan Schonfeld, Ben Sender, Jonathan D. Tang, Emma Tsurkov, Austin van Loon, Onur Varol, Xiafei Wang, Zhi Wang, Julia Wang, Flora Wang, Samantha Weissman, Kirstie Whitaker, Maria K Wolters, Wei Lee Woon, James Wu, Catherine Wu, Kengran Yang, Jingwen Yin, Bingyu Zhao, Chenyun Zhu, Jeanne Brooks-Gunn, Barbara E. Engelhardt, Moritz Hardt, Dean Knox, Karen Levy, Arvind Narayanan, Brandon M. Stewart, Duncan J. Watts, and Sara McLanahan. “Measuring the predictability of life outcomes with a scientific mass collaboration.” Proceedings of the National Academy of Sciences. Mar 2020, 201915006; DOI: 10.1073/pnas.1915006117

Briana Mezuk, Tomohiro Ko, Viktoryia A. Kalesnikava, **David Jurgens**. “Suicide among older adults living in or transitioning to residential long-term care: 2003–2015” *JAMA Network Open* 2019;2(6):e195627

Ashwin Rajadesingan, Ramaswami Mahalingam, and **David Jurgens**. “Smart, Responsible, and Upper Caste Only: Measuring Caste Attitudes through Large-Scale Analysis of Matrimonial Profiles.” Proceedings of the AAAI International Conference on Web and Social Media (ICWSM), 2019.² **Best Paper Award**.

David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. 2018. “Citation Classification for Behavioral Analysis of a Scientific Field.” Transactions of the Association for Computational Linguistics (TACL).

Rob Voigt, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Camilla M. Griffiths, **David Jurgens**, Dan Jurafsky, and Jennifer L. Eberhardt. 2017. “Language from police body camera footage shows racial disparities in officer respect.” Proceedings of the National Academy of Science (PNAS). Winner of the **Cialdini Prize** and **Cozzarelli Prize**

Osman Başkaya and **David Jurgens**. Word Sense Induction and Disambiguation Rivaling Supervised Methods. *Journal of Artificial Intelligence Research* (JAIR). 55(1) pp 1025-1058. 2016.

David Jurgens, Mohammad Taher Pilehvar, and Roberto Navigli. Cross Level Semantic Similarity: An Evaluation Framework for Universal Measures of Similarity. *Journal of Language Resources and Evaluation*. 2015

David Jurgens and Roberto Navigli. It’s All Fun and Games until Someone Annotates: Video Games with a Purpose for Linguistic Annotation. *Transactions of the Association for Computational Linguistics* (TACL). 2014.

Peer-Reviewed Conference Proceedings

Aparna Ananthasubramaniam, Yufei Zhu, **David Jurgens**, Daniel Romero. “The Role of Network and Identity in the Diffusion of Hashtags”. Proceedings of the Web Conference. 2025.

²Beginning in 2019, ICWSM became a journal with a full R&R process and rolling submission

Huaman Sun, Jiaxin Pei, Minje Choi, and **David Jurgens**. “Sociodemographic Prompting is Not Yet an Effective Approach for Simulating Subjective Judgments with LLMs” Proceedings of the 2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). 2025.

Yinuo Xu, Hong Chen, Sushrita Rakshit, Aparna Ananthasubramaniam, Omkar Yadav, Mingqian Zheng, Michael Jiang, Lechen Zhang, Bowen Yi, Kenan Alkiek, Abraham Israeli, Bangzhao Shu, Hua Shen, Jiaxin Pei, Haotian Zhang, Miriam Schirmer, and **David Jurgens**. “Causally Modeling the Linguistic and Social Factors that Predict Email Response” Proceedings of the 2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). 2025.

Chengxing Xie, Canyu Chen, Feiran Jia, Ziyu Ye, Kai Shu, Adel Bibi, Ziniu Hu, **David Jurgens**, James Evans, Philip Torr, Bernard Ghanem, and Guohao Li. “Can Large Language Model Agents Simulate Human Trust Behaviors?” Proceedings of the 37th Conference on Advances in Neural Information Processing Systems (NeurIPS). 2024.

Chan Young Park, Shuyue Stella Li, Hayoung Jung, Svitlana Volkova, Tanushree Mitra, **David Jurgens**, Yulia Tsvetkov. “ValueScope: Unveiling Implicit Norms and Values via Return Potential Model of Social Interactions” Findings of EMNLP. 2024.

Miriam Schirmer, Tobias Leemann, Gjergji Kasneci, Jürgen Pfeffer, **David Jurgens**. “The Language of Trauma: Modeling Traumatic Event Descriptions Across Domains with Explainable AI” Findings of EMNLP. 2024.

Mingqian Zheng, Jiaxin Pei, Moontae Lee, Lajanugen Logeswaran, and **David Jurgens**. “Is ‘A Helpful Assistant’ the Best Role for Large Language Models? A Systematic Evaluation of Social Roles in System Prompts” Findings of EMNLP. 2024.

Tong Lin*, Jason Yan*, **David Jurgens**, and Sabina Tomkins. “Tab2Text - A framework for deep learning with tabular data” Findings of EMNLP. 2024.

Bangzhao Shu*, Lechen Zhang*, Minje Choi, Lavinia Dunagan, Lajanugen Logeswaran, Moontae Lee, Dallas Card, and **David Jurgens**. “You don’t need a personality test to know these models are unreliable: Assessing the Reliability of Large Language Models on Psychometric Instruments.” Proceedings of the 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). 2024.

Social Meme-ing: Measuring Linguistic Variation in Memes Naitian Zhou, **David Jurgens**, and David Bamman. Proceedings of the 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). 2024.

Jiamin Yang and **David Jurgens**. “Modeling Empathetic Alignment in Conversation.” Proceedings of the 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). 2024

Xi Chen, Scott A. Hale, **David Jurgens**, Mattia Samory, and Przemyslaw. “A Multilingual Similarity Dataset for News Article Frames.” Proceedings of the International AAAI Conference on Web and Social Media (ICWSM) Posters, Demos, and Datasets Track. 2024

Xi Chen, Scott A. Hale, **David Jurgens**, Mattia Samory, Ethan Zuckerman, Przemyslaw, Adam Grabowicz. “Global News Synchrony During the Start of the COVID-19 Pandemic.” Proceedings of the 2024 Web Conference. 2024.

Sungjin Nam, Kevyn Collins-Thompson, **David Jurgens** and Xin Tong. “Finding Educationally Supportive Contexts for Vocabulary Learning with Attention-Based Models.” Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING). 2024

Minje Choi*, Jiaxin Pei*, Sagar Kumar, Chang Shu, and **David Jurgens**. “Do LLMs Understand Social Knowledge? Evaluating the Sociability of Large Language Models with SocKET Benchmark.” Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2023.

Ben Litterer, **David Jurgens**, and Dallas Card. “When it Rains, it Pours: Modeling Media Storms and the News Ecosystem.” Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2023.

David Jurgens,* Agrima Seth,* Jackson Sargent,[†] Athena Aghighi,[†] and Michael Geraci.[†] “Your spouse needs professional help: Determining the Contextual Appropriateness of Messages through Modeling Social Relationships.” Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL). 2023.

Dustin Wright, Jiaxin Pei, **David Jurgens**, and Isabelle Augenstein. “Modeling Information Change in Science Communication with Semantically Matched Paraphrases.” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2022.

Allison Claire Lahnala, Charles Welch, **David Jurgens**, and Lucie Flek. “A Critical Reflection and Forward Perspective on Empathy and Natural Language Processing.” Proceedings of the Findings of Empirical Methods in Natural Language Processing (EMNLP Findings). 2022.

Jiaxin Pei, Aparna Kamakshi Ananthasubramaniam, Xingyao Wang, Naitian Zhou, Apostolos Dedeloudis, Jackson Sargent, and **David Jurgens**. “POTATO: The Portable Text Annotation Tool.” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP): Systems Demonstrations. 2022.

Anne Lauscher, Brandon Ko, Bailey Kuhl, Sophie Johnson, Arman Cohan, **David Jurgens**, Kyle Lo. “MultiCite: Modeling realistic citations requires moving beyond the single-sentence single-label setting.” Proceedings of the 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). 2022.

Jian Zhu, Cong Zhang, and **David Jurgens**. “ByT5 model for massively multilingual grapheme-to-phoneme conversion” INTERSPEECH 2022.

Kenan Alkiek, Bohan Zhang, and **David Jurgens**. “Comparing Apples and Oranges: Recognizing Political Heterogeneity on Reddit and Its Implications for Behavioral Analysis.” Proceedings of the Findings of the Annual Meeting of the Association for Computational Linguistics (Findings of ACL), 2022.

Jian Zhu, Cong Zhang, and **David Jurgens**. “Phone-to-audio alignment without text: A Semi-supervised Approach.” Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022.

Sky Wang and **David Jurgens**. “Using Sociolinguistic Variables to Reveal Changing Attitudes Towards Sexuality and Gender.” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.

Jian Zhu and **David Jurgens**. “Learning Idiolectal Style Variations in Online Register.” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.

Jiaxin Pei and **David Jurgens**. “Measuring Sentence-Level and Aspect-Level Certainty in Science Communications” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.

Chan Young Park, Julia Mendelsohn, Karthik Radhakrishnan, Kinjal Jain, Tushar Kanakagiri, **David Jurgens** and Yulia Tsvetkov. “Detecting Community Norm Violations in Online Conversations.” Proceedings of the Findings of the Conference on Empirical Methods in Natural Language Processing (Findings of EMNLP), 2021.

Xingyao Wang and **David Jurgens**. “An Animated Picture Says at Least a Thousand Words: Selecting Gif-based Replies in Multimodal Dialog.” Proceedings of the Findings of the Conference on Empirical Methods in Natural Language Processing (Findings of EMNLP), 2021.

Julia Mendelsohn, Ceren Budak, and **David Jurgens**. “Modeling Framing in Immigration Discourse on Social Media.” Proceedings of the North American Meeting of the Association for Computational Linguistics (NAACL), 2021.

Jian Zhu and **David Jurgens**. “The structure of online social networks modulates the rate of lexical change.” Proceedings of the North American Meeting of the Association for Computational Linguistics (NAACL), 2021.

Jiajun Bao,* Junjie Wu,* Yiming Zhang,* Eshwar Chandrasekharan, and **David Jurgens**. “Conversations Gone Alright: Quantifying and Predicting Prosocial Outcomes in Online Conversations.” Proceedings of the Web Conference (WWW/WebConf), 2021.

“Still out there: Modeling and Identifying Russian Troll Accounts on Twitter.” Jane Im, Eshwar Chandrasekharan, Jackson Sargent, Paige Lighthammer, Taylor Denby, Ankit Bhargava, Libby Hemphill, **David Jurgens**, Eric Gilbert. [\[link\]](#) Proceedings of Web Science, 2020. **Best Paper Runner Up**.

Jiaxin Pei and **David Jurgens**. “Quantifying Intimacy In Language” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.

Naitian Zhou and **David Jurgens**. “Condolence and Empathy in Online Communities” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.

Luke Breitfeller, Emily Ahn, **David Jurgens**, and Yulia Tsvetkov. “Finding Microaggressions in the Wild: A Case for Locating Elusive Phenomena in Social Media Posts.” Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019.

Meixing Dong, **David Jurgens**, Carmen Banea and Rada Mihalcea. “Perceptions of social roles across cultures.” Proceedings of Social Informatics (SocInfo), 2019. **Nominated for Best Paper**.

Innocent Ndubuisi-Obi*, Sayan Ghosh*, **David Jurgens**. “Wetin dey with these comments? Modeling Sociolinguistic Factors Affecting Code-switching Behavior in Nigerian Online Discussions.” Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2019

David Jurgens, Eshwar Chandrasekharan, Libby Hemphill. “A Just and Comprehensive Strategy for Using NLP to Address Online Abuse.” Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), 2019

Zijian Wang, Scott Hale, David Ifeoluwa Adelani, Przemyslaw Grabowicz, Timo Hartmann, Fabian Flöck and **David Jurgens**³ “Demographic Inference and Representative Population Estimates from Multilingual Social Media Data.” Proceedings of the Web Conference (WWW), 2019

Tiago Cunha, **David Jurgens**, Chenhao Tan and Daniel Romero. “Are All Successful Communities Alike? Characterizing and Predicting the Success of Online Communities.” Proceedings of the Web Conference (WWW/WebConf), 2019

Zijian Wang and **David Jurgens**. “It’s going to be okay: Measuring Access to Support in Online Communities.” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018

David Jurgens, James McCorriston, and Derek Ruths. 2017. An Analysis of Individuals’ Behavior Change in Online Groups. Proceedings of the 9th International Conference on Social Informatics (SocInfo).

David Jurgens, Yulia Tsvetkov, and Dan Jurafsky. 2017. Writer Profiling Without the Writer’s Text. Proceedings of the 9th International Conference on Social Informatics (SocInfo).

David Jurgens, Yulia Tsvetkov, and Dan Jurafsky. 2017. Incorporating Dialectal Variability for Socially Equitable Language Identification. In Proceedings of the Annual Meeting for the Association for Computational Linguistics (ACL).

Edward Newell*, **David Jurgens ***, Hardik Vala, Jad Sassine, Caitrin Armstrong, Derek Ruths and Haji Mohammad Saleem. 2016. Social Migration in the Digital Age: A Case Study on Reddit During A Period of Community Unrest. Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM).

Hardik Vala, **David Jurgens**, Andrew Piper, and Derek Ruths. 2015. Mr. Bennet, his coachman, and the Archbishop walk into a bar but only one of them gets recognized: On The Difficulty of Detecting Characters in Literary Texts. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).

Silvia Neculescu, Sara Mendes, **David Jurgens**, Núria Bel, and Roberto Navigli. 2015. Reading Between the Lines: Overcoming Data Sparsity for Accurate Classification of Lexical Relationships. Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics (*SEM).

³Corresponding senior author

David Jurgens and Mohammad Taher Pilehvar. 2015. Reserating the awesometastic: An automatic extension of the WordNet taxonomy for novel terms. In Proceeding of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL-HLT).

David Jurgens, Tyler Finethy, James McCorriston, Yi Tian Xu, and Derek Ruths. 2015. Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. In Proceedings of the 9th International Conference on Weblogs and Social Media (ICWSM).

David Jurgens, James McCorriston, and Derek Ruths. 2015. An Analysis of Exercising Behavior in Online Populations. In Proceedings of the 9th International Conference on Weblogs and Social Media (ICWSM).

James McCorriston, **David Jurgens**, and Derek Ruths. 2015. Organizations are Users Too: Characterizing and Detecting the Presence of Organizations on Twitter. In Proceedings of the 9th International Conference on Weblogs and Social Media (ICWSM).

Ryan Compton, **David Jurgens**, and David Allan. 2014. Geotagging One Hundred Million Twitter Accounts with Total Variation Minimization. In Proceedings of the IEEE International Conference on Big Data.

Daniele Vannella, **David Jurgens**, Daniele Scarfini, Domenico Toscani, and Roberto Navigli. 2014. Validating and Extending Semantic Knowledge Bases using Video Games with a Purpose. In Proceedings of the Annual Meeting for the Association for Computational Linguistics (ACL).

Mohammad T. Pilehvar, **David Jurgens**, and Roberto Navigli. 2013. Align, Disambiguate and Walk: A Unified Approach for Measuring Semantic Similarity. In Proceedings of the Annual Meeting for the Association for Computational Linguistics (ACL). **Best Paper Nomination**.

David Jurgens. 2013. That’s what friends are for: Inferring location in online communities based on social relationships. In Proceedings of the 7th International Conference on Weblogs and Social Media (ICWSM).

David Jurgens. 2013. Embracing Ambiguity: A Comparison of Annotation Methodologies for Crowdsourcing Word Sense Labels. In Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL).

David Jurgens. 2012. An Evaluation of Graded Sense Disambiguation using Word Sense Induction. In Proceedings of *SEM, the First Joint Conference on Lexical and Computational Semantics.

David Jurgens and Tsai-Ching Lu. 2012. Temporal Motifs Reveal the Dynamics of Editor Interactions in Wikipedia. In Proceedings of the 6th International Conference on Weblogs and Social Media (ICWSM).

David Jurgens and Keith Stevens. 2010. The S-Space Package: An Open-Source Framework for Word Space Algorithms. Proceedings of the ACL System Demonstrations.

Lightly Peer-Reviewed Conference Proceedings

Rob Voigt, **David Jurgens**, Vinodkumar Prabhakaran, Dan Jurafsky, and Yulia Tsvetkov. 2018. “RtGender: A Corpus of Responses to Gender for Studying Gender Bias.” Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC).

Hardik Vala, Stefan Dimitrov, **David Jurgens**, Andrew Piper and Derek Ruths. 2016. Annotating Characters in Literary Corpora: A Scheme, the CHARLES Tool, and an Annotated Novel. Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC).

David Jurgens. 2014. An analysis of ambiguity in word sense annotations. Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC).

Workshop and Shared Task Papers⁴

⁴Shared Task papers are typically heavily peer-reviewed during the task proposal phase, when authors submit detailed descriptions of a potential task to be included in the workshop (~20-30% Acceptance Rate); the corresponding shared task paper details the task, data construction, participants, and summarizes the findings but are lightly reviewed.

Jiaxin Pei and **David Jurgens**. “When Do Annotator Demographics Matter? Measuring The Influence of Annotator Demographics with the POPQUORN Dataset.” Proceedings of the 17th Linguistic Annotation Workshop (LAW-XVII) at ACL. 2023.

Aparna Ananthasubramaniam, Hong Chen, Jason Yan, Kenan Alkiek, Jiaxin Pei, Agrima Seth, Lavinia Dunagan, Minje Choi, Benjamin Litterer, and **David Jurgens**. “Exploring Linguistic Style Matching in Online Communities: The Role of Social Context and Conversation Dynamics.” Proceedings of the 1st Workshop on Social Influence in Conversations (SICon) at ACL. 2023. **Best Paper**

Christina Lu and **David Jurgens**. “The subtle language of exclusion: Identifying the Toxic Speech of Trans-exclusionary Radical Feminists.” Proceedings of the 6th Workshop on Online Abuse and Harms (WOAH). 2022.

Xi Chen, Ali Zeynali, Chico Camargo, Fabian Flöck, Devin Gaffney, Przemyslaw Grabowicz, Scott Hale, **David Jurgens**, and Mattia Samory. 2022. SemEval-2022 Task 8: Multilingual news article similarity. In Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval). 2022.

Sayan Ghosh, Dylan Baker, **David Jurgens**, and Vinodkumar Prabhakaran. “Detecting Cross-Geographic Biases in Toxicity Modeling on Social Media.” Proceedings of the 7th Workshop on Noisy User-generated Text (W-NUT). 2021. **Best Paper**.

Huiyang Ding and **David Jurgens**. 2021. HamiltonDinggg at SemEval-2021 Task 5: Investigating Toxic Span Detection using RoBERTa Pre-training. Proceedings of the 2021 Workshop on Semantic Evaluation.

David Jurgens. “Learning PyTorch Through A Neural Dependency Parsing Exercise.” Proceedings of the Fifth Workshop on Teaching NLP, 2021.

David Jurgens. “Learning about Word Vector Representations and Deep Learning through Implementing Word2vec.” Proceedings of the Fifth Workshop on Teaching NLP, 2021.

David Jurgens and Mohammad Taher Pilehvar. 2016. SemEval-2016 Task 14: Semantic Taxonomy Enrichment. In Proceedings of the 10th Workshop on Semantic Evaluation.

Jussi Karlgren, Jimmy Callin, Kevyn Collins-Thompson, Amaru Cuba Gyllensten, Ariel Ekgren, **David Jurgens**, Anna Korhonen, Fredrik Olsson, Magnus Sahlgren, and Hinrich Schütze. 2015. Evaluating learning language representations. Proceedings of Conference and Labs of Evaluation Forum (CLEF).

David Jurgens, Tyler Finethy, Caitrin Armstrong, and Derek Ruths. 2015. Everyone’s Invited: A New Paradigm For Evaluation on Non-transferable Datasets. In Proceedings of ICWSM Workshop on Standards and Practices in Large-Scale Social Media Research.

David Jurgens, Stefan Dimitrov, and Derek Ruths. 2014. Twitter users #CodeSwitch hashtags! #MoltoImportante #wow #현. Proceedings of the EMNLP Workshop on Code Switching.

David Jurgens, Mohammad T. Pilehvar, and Roberto Navigli. 2014. SemEval-2014 Task 3: Cross-Level Semantic Similarity. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014).

Roberto Navigli, **David Jurgens** and Daniele Vannella. 2013. SemEval-2013 Task 12: Multilingual Word Sense Disambiguation. In Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013).

David Jurgens and Ioannis Klapaftis. 2013. SemEval-2013 Task 13: Word Sense Induction for Graded and Non-Graded Senses. In Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013).

Veronika Strnadova, **David Jurgens**, Tsai-Ching Lu. 2013. Characterizing Online Discussions in Microblogs Using Network Analysis. In Proceedings of the AAAI Spring Symposium on Analyzing Microtext.

David Jurgens, Saif M. Mohammad, Peter D. Turney, and Keith J. Holyoak. 2012. SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. In Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012).

David Jurgens. 2011. Word sense induction by community detection. In Proceedings of Sixth ACL Workshop on Graph-based Methods for Natural Language Processing (TextGraphs-6).

David Jurgens and Keith Stevens. 2011. Measuring the impact of sense similarity on word sense induction. In Proceedings of the First Workshop on Unsupervised Learning in NLP (UNSUP-2011), pages 113–123.

David Jurgens and Keith Stevens. 2010. Capturing Nonlinear Structure in Word Spaces Through Dimensionality Reduction. Proceedings of the ACL 2010 Workshop on GEometrical Models of Natural Language Semantics (GEMS).

David Jurgens and Keith Stevens. 2010. HERMIT: Flexible Clustering for the SemEval-2 WSI Task. In Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval 2010).

David Jurgens and Keith Stevens. 2009. Event Detection in Blogs using Temporal Random Indexing. In Proceedings of the International Workshop on Events on Emerging Text Types (eETTs), pages 21–28.

Robert Pless and **David Jurgens**. Road extraction from motion cues in aerial video. In Proceedings of the ACM Conference on Geographic Information Systems, pages 31–38, 2004.

Abstracts

Benjamin Litterer, **David Jurgens**, and Dallas Card “Adventures in Audio: An Open and Comprehensive Dataset of Podcasts for CSS Research” Abstract at the International Conference on Computational Social Science (IC2S2), 2024.

Xi Chen, Scott A. Hale, **David Jurgens**, Mattia Samory, Ethan Zuckerman, Przemyslaw A. Grabowicz “Global News Synchrony and Diversity During the Start of the COVID-19 Pandemic.” Abstract at the International Conference on Computational Social Science (IC2S2), 2024.

Minje Choi, Daniel Romero, and **David Jurgens** “Profile Update: The Effects of Identity Disclosure on Network Connections and Language.” Abstract at the International Conference on Computational Social Science (IC2S2), 2024.

Aparna Ananthasubramaniam, Yufei Louise Zhu, **David Jurgens**, and Daniel Romero “Roles of Network and Identity in Hashtag Diffusion” Abstract at the International Conference on Computational Social Science (IC2S2), 2024.

Aparna Ananthasubramaniam, Daniel Romero, and **David Jurgens**. “Macroeconomic Context Moderates Socioeconomic Determinants of Suicide” Abstract at the International Conference on Computational Social Science (IC2S2), 2024.

Sky Wang and **David Jurgens**. “Analyzing Changing Attitudes Towards Sexuality and Gender Using Lexical Variables in Social Media” Abstract at the New Ways of Analyzing Variation (N WAV) conference, 2022.

Xi Chen, Ali Zeynali, Chico Camargo, Fabian Flöck, Devin Gaffney, Przemyslaw Grabowicz, Scott Hale, **David Jurgens**, and Mattia Samory. “Multilingual News Article Similarity.” Abstract at the International Conference on Computational Social Science (IC2S2), 2022.

Leopele Raabe, Misha Teplitskiy, and **David Jurgens**. “Measuring Cross-partisan Exposure through Physical Mobility Patterns.” Abstract at the International Conference on Computational Social Science (IC2S2), 2022.

Jiaxin Pei and **David Jurgens**. “Do Journalists Overstate Science? Findings from Computational Modeling of Scientific Uncertainty.” Abstract at the International Conference on Computational Social Science (IC2S2), 2022.

Aparna Ananthasubramaniam, **David Jurgens**, and Daniel Romero. “Here, There, or Everywhere: Networks and Identity Drive Geographic Properties of the Diffusion of Linguistic Innovation.” Abstract at the International Conference on Computational Social Science (IC2S2), 2022.

Minje Choi, **David Jurgens**, and Daniel Romero. “Analyzing the Engagement of Social Relationships During Life Event Shocks in Social Media.” Abstract at the International Conference on Computational Social Science (IC2S2), 2022.

Julia Mendelsohn, Sayan Ghosh, Ceren Budak and **David Jurgens**. “Bridging Nations: Quantifying the Influence of Multilinguals in the European Twitter Network” Abstract at the International Conference on Computational Social Science (IC2S2), 2021.

Julia Mendelsohn, Ceren Budak, and **David Jurgens**. “Modeling the framing of immigration on social media” Abstract at the International Conference on Computational Social Science (IC2S2), 2021.

Jiaxin Pei and **David Jurgens**. “Computational Modeling of Intimacy in Language and Social Norms in Interpersonal Communications” Abstract at the International Conference on Computational Social Science (IC2S2), 2021.

Hao Peng, Misha Teplitskiy, and **David Jurgens** “Say My Name, Say My Name: Author Mentions in Science Reporting Reveal Wide-Spread Ethnic Bias” Abstract at the International Conference on Computational Social Science (IC2S2), 2020.

Ashwin Rajadesingan, Ramaswami Mahalingam, and **David Jurgens** “Measuring Caste Discrimination using Matrimonial Profiles” Abstract at the International Conference on Computational Social Science (IC2S2), 2020.

Minje Choi, Ceren Budak, Daniel Romero, and **David Jurgens** “Interpersonal relationships and its effect on communication behavior in Twitter” Abstract at the International Conference on Computational Social Science (IC2S2), 2020.

Jiajun Bao, Junjie Wu, Yiming Zhang, Eshwar Chandrasekharan, and **David Jurgens** “Measuring and Forecasting Prosocial Behavior in Social Media” Abstract at the International Conference on Computational Social Science (IC2S2), 2020.

Aparna Ananthasubramaniam, **David Jurgens**, and Daniel Romero. “On the Localization of New Words on Twitter” Abstract at the International Conference on Computational Social Science (IC2S2), 2020.

Zijian Wang, Scott Hale, David Ifeoluwa Adelani, Przemyslaw Grabowicz, Timo Hartmann, Fabian Flöck and **David Jurgens** “Demographic Inference and Representative Population Estimates from Multilingual Social Media Data.” Abstract at the International Conference on Computational Social Science (IC2S2), 2019.

David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. 2018. “Citation Classification for Behavioral Analysis of a Scientific Field.” Abstract at the International Conference on Computational Social Science (IC2S2), 2018.

David Ifeoluwa Adelani, Fabian Flöck, Przemyslaw Grabowicz, Scott Hale, Timo Hartman and **David Jurgens** “Agendas on Social Media: Inferring Policy Attention from Non-representative Data” Abstract at the International Conference on Computational Social Science (IC2S2), 2018.

David Jurgens and Tsai-Ching Lu. 2012. Friends, Enemies, and Lovers: Detecting Communities in Networks Where Relationships Matter. In Proceedings of Web Science. ACM.

Manuscripts and Preprints

Anna Wegmann, Dong Nguyen, and **David Jurgens**. “Tokenization is Sensitive to Language Variation.” [\[link\]](#)

Dustin Wright, Zain Muhammad Mujahid, Lu Wang, Isabelle Augenstein, and **David Jurgens**. “Unstructured Evidence Attribution for Long Context Query Focused Summarization.” [\[link\]](#)

Shivani Kumar and **David Jurgens**. “Are Rules Meant to be Broken? Understanding Multilingual Moral Reasoning as a Computational Pipeline with UniMoral.” [\[link\]](#)

Allison Lahnala, Charlie Welch, **David Jurgens**, Lucie Flek. “The Muddy Waters of Modeling Empathy in Language: The Practical Impacts of Theoretical Constructs.” [\[link\]](#)

Mapping the Podcast Ecosystem with the Structured Podcast Research Corpus Benjamin Litterer, **David Jurgens**, Dallas Card [\[link\]](#)

Lechen Zhang, Tolga Ergen, Lajanugen Logeswaran, Moontae Lee, and **David Jurgens**. “SPRIG: Improving Large Language Model Performance by System Prompt Optimization.” [\[link\]](#)

Johnathan Ivey, [Shivani Kumar](#), [Jiayu Liu](#), [Hua Shen](#), [Sushrita Rakshit](#), [Rohan Raju](#), [Haotian Zhang](#), [Aparna Ananthasubramaniam](#), [Junghwan Kim](#), [Bowen Yi](#), [Dustin Wright](#), [Abraham Israeli](#), [Anders Giovanni Møller](#), [Lechen Zhang](#), and **David Jurgens**. “Real or Robotic? Assessing Whether LLMs Accurately Simulate Qualities of Human Responses in Dialogue.” [\[link\]](#)

[Abraham Israeli](#), **David Jurgens**, and Daniel Romero. “A Test of Time: Predicting the Sustainable Success of Online Collaboration in Wikipedia.” [\[link\]](#)

[Hua Shen](#), [Tiffany Knearem](#), [Reshmi Ghosh](#), [Kenan Alkiek](#), [Kundan Krishna](#), [Yachuan Liu](#), [Ziqiao Ma](#), [Savvas Petridis](#), [Yi-Hao Peng](#), [Li Qiwei](#), [Sushrita Rakshit](#), [Chenglei Si](#), [Yutong Xie](#), [Jeffrey P Bigham](#), [Frank Bentley](#), [Joyce Chai](#), [Zachary Lipton](#), [Qiaozhu Mei](#), [Rada Mihalcea](#), [Michael Terry](#), [Diyi Yang](#), [Meredith Ringel Morris](#), [Paul Resnick](#), **David Jurgens**. “Towards Bidirectional Human-AI Alignment: A Systematic Review for Clarifications, Framework, and Future Directions” [\[link\]](#)

[Hong Chen](#), [Christopher I Rider](#), **David Jurgens**, [Misha Teplitskiy](#). “Geographical disparities in navigating rejection in science drive disparities in its file drawer” [\[link\]](#)

[Sungjin Nam](#), **David Jurgens**, and [Kevyn Collins-Thompson](#) “Attention-based Learning Models for Predicting Contextual Informativeness.” [\[link\]](#).

TECHNICAL TUTORIALS [Jack Grieve](#) (University of Birmingham) and **David Jurgens** (University of Michigan) “[Mapping word frequencies on Twitter using R and Python](#)” at New Ways of Analyzing Variation (NWAV) 48. October 10, 2019.

[Jack Grieve](#) (University of Birmingham), [Dirk Hovy](#) (Bocconi University), [David Jurgens](#) (University of Michigan), [Tyler Kendall](#) (University of Oregon), [Dong Nguyen](#) (Alan Turing Institute), [James Stanford](#) (Dartmouth College), [Meghan Sumner](#) (Stanford University), [Rachael Tatman](#) (Kaggle). “Computational Sociolinguistics” at New Ways of Analyzing Variation (NWAV) 47. October 18, 2018. (50+ participants)

David Jurgens and [Mohammad Taher Pilehvar](#). Semantic Similarity Frontiers: From Concepts to Documents. Presented at the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015. (Over 155 registered participants)

[Roberto Navigli](#) and **David Jurgens**. Multilingual Semantic Processing. Presented at the International Conference on Language Resources and Evaluation (LREC), 2014. (33 registered participants)

INVITED TALKS AND PANELS “What exactly is alignment? A big-picture view of the challenges and opportunities of AI alignment” **Keynote** at the Getting Aligned On AI Alignment Workshop. University of Maryland, College Park. 1 November 2024.

Panelist at the New Directions: Bridging Natural Language Processing (NLP) and Survey Research (SurvAI) Workshop. University of Maryland, College Park. 7 October 2024.

“Do foundation models have personalities? The risks and opportunities for model alignment and personification” AI24 Conference hosted by the Lammar Institute. Dortmund, Germany. 5 September 2024.

“The Challenges and Opportunities of LLM Alignment and Personalization” **Keynote** at the Workshop on the Frontiers of AI in Business and Society. University of Illinois, Chicago. 10 May 2024.

“Towards a Deep Understanding of Social Language and Behavior through Deep Learning and Natural Language Processing”. Seminar for the Social Psychology program. University of Michigan. 17 April 2024.

“Assessing the Social Capacity of Large Language Models.” University of Oxford. 30 Nov. 2023.

“The Social Dimensions of Communication: How Context Shapes Language Use and Interpretation” **Keynote** at the Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA) at ACL 2023. 14 July 2023.

“Beyond the Individual: Modeling Interpersonal Relationships in NLP Studies of Communication” University of Stuttgart. 1 March 2023.

“Venturing Off the NLP Map into Interdisciplinary Lands” **Keynote** at the Share Stories and Lessons Learned workshop at EMNLP 2022. 8 December 2022.

“The challenges and opportunities of defining what is political in social media.” **Keynote** at the Workshop on Computational Linguistics for Political Text Analysis. 12 September 2022.

“It’s not who you know, it’s how you know them: Modeling and Understanding the Role of Social Relationships in Social Media Communication” **Keynote** at the W-NUT workshop at COLING 2022. 16 October 2022.

“Sometimes good things happen on social media but sometimes you have to make them happen.” University of California, Santa Barbara. 24 January 2022.

“Modeling the Framing of Immigration Discourse and its Effects in Social Media.” HRL Laboratories. 18 March 2021.

“Modeling the Framing of Marginalized/Politicized People in Social and News Media.” GESIS Leibniz Institut für Sozialwissenschaften. 9 March 2021.

“Social Interpretations of Interpersonal Communication.” **Keynote** at the PEOPLES workshop at COLING 2020. 13 December 2020.

“Social Interpretations of Interpersonal Communication.” University of Mannheim. 1 December 2020.

“News flash! Author mentions in science news reveal wide-spread ethnic bias.” Michigan Center for Urban African American Aging Research (MCUAAAR). 18 November 2020

“Citation Classification for Behavioral Analysis of a Scientific Field.” **Keynote** at the 8th International Workshop on Mining Scientific Publications. 5 August 2020.

“Putting a Face on Science: Analyzing Author Mentions in Science Journalism Reveal Wide-Spread Ethnic Bias.” SciNLP workshop at AKBC2020. 24 June 2020.

“Large-scale population measurements from social media.” Adansonia Conference at Bocconi University, Milan Italy. 30 May 2019

“Uncovering large scale social processes through computational linguistic analysis.” UCLA Computational Sociology seminar. 28 February 2019.

“Demographic Inference in Social Media for Large-Scale Sociolinguistic Analysis.” Lorentz Center in Leiden, Netherlands. 6 November, 2019.

“Large-scale Social Studies using NLP.” Language and Information Technologies Laboratory at the University of Michigan. 3 October, 2017.

“Citation Classification for Behavioral Analysis of a Scientific Field.” University of California, Berkeley. 14 November 2016.

“Observing cultural exchange across linguistic and geographic boundaries in social media.” The workshop on Challenges for Data-driven Natural Language Analysis beyond Standard Data at the Institut für Maschinelle Sprachverarbeitung. 14 October 2016.

“Observing cultural exchange across linguistic and geographic boundaries in social media.” Stanford Data Science Initiative. 11 October 2016.

“Linguistic Annotation Using Video Games with a Purpose.” Information Sciences Institute at the University of Southern California. 4 March 2016.

“Towards a Universal Method for Measuring Semantic Textual Similarity.” University of Stuttgart. 3 July 2014.

“Productively Slacking Off: Building video games with a purpose for large-scale linguistic annotation.” Washington University in St. Louis. 23 November 2015.

TEACHING

University of Michigan, School of Information

- SI 311-110, How to use AI Effectively (Fall 2024). Enrollment ~ 50
- SI 330, Data Manipulation (Winter 2023). Enrollment ~ 75
- SI 630, Natural Language Processing: Algorithms and People (Winter 2018–2024). Enrollments ~ 75 –125
- SI 650, Information Retrieval (Fall 2020–2023). Enrollments ~ 50 –75
- SI 671/721, Data Mining: Methods and Applications (Fall 2017, 2018). Enrollments ~ 75
- SI 710, Doctoral Seminar in Computational Sociolinguistics (Fall 2019). Enrollment of 15
- SIADS 655,[†] Applied Natural Language Processing (Fall 2020, Winter 2021). Enrollments ~ 75

[†] denotes 1-credit courses taught as a part of the Michigan Applied Data Science degree.

University of Michigan, Department of Electrical Engineering and Computer Science

- EECS 549, Information Retrieval (Fall 2020–2023). Enrollments ~ 50 –75
- EECS 595, Natural Language Processing (Winter 2024). Enrollments ~ 50 –75

ADVISING AND MENTORING

Graduate students

Kenan Alkiek (PhD, School of Information; Fall 2022–*present*), Co-advised with V.G. Vinod Vydiswaran

Hong Chen (PhD, School of Information; Fall 2022–*present*), Co-advised with Misha Teplitskiy

Minje Choi (PhD, School of Information; Fall 2018–Summer 2023), Co-advised with Daniel Romero → Postdoc at GaTech.

Junghwan Kim (PhD, Computer Science; Summer 2024–*present*)

Eleanor Lin (PhD, Computer Science; Fall 2024–*present*)

Ben Litterer (PhD, School of Information; Fall 2022–*present*), Co-advised with Dallas Card

Jiaxin Pei (PhD, School of Information; Fall 2018–*present*)

Nancy Xu (PhD, School of Information; Fall 2023–*present*)

Jian Zhu (PhD Linguistics; Fall 2020–Summer 2022) → Incoming Faculty at the University of British Columbia.

Doctoral Dissertation Committees

Afshin Rahimi, Computer Science, 2018 (University of Melbourne)

Steve Wilson, Computer Science, 2019 (U-M)

Sungjin Nam, School of Information, 2020 (U-M)

Ian Stewart, School of Interactive Computing, 2020 (Georgia Tech)

Prashant Rajaram, Business, 2021 (U-M)

Susannah Chandhok, Social Psychology, 2021 (U-M)

Michael Miller Yoder, Language Technology Institute, 2021 (Carnegie Mellon University)

Qinlan Shen, Language Technology Institute, 2021 (Carnegie Mellon University)

MeiXing Dong, Computer Science, 2021 (U-M)
 James Ravenscroft, Urban Science, 2022 (University of Warwick, UK)
 Terne Jakobsen, Computer Science, 2023 (University of Copenhagen)
 Michael Ion, School of Education, 2024 (U-M)
 Amrit Romana, Computer Science, 2024 (U-M)
 Chan Park, Language Technology Institute, *expected 2024* (Carnegie Mellon University),
 Andrew McCrab, Computer Science, *expected 2024* (U-M) Jason Godfrey, School of
 Education, *expected 2024* (U-M) Annalise Paulson, School of Education, *expected 2025*
 (U-M)

Awards Won By My Students

I actively mentor students to help them shape their research portfolios. Multiple students that I have mentored have been subsequently recognized for their research through awards and prestigious fellowships. Students are listed alphabetically.

- Lavinia Dunagan: NSF GRFP
- Kenan Alkiek: Theresa Noel Urban Blaurock Research Award
- Agrima Seth: Snap Fellowship Honorable Mention
- Julia Mendelsohn: NSF GRFP Honorable Mention, Snap Fellowship Honorable Mention, Meta (Facebook) Fellowship Finalist, Google Research Fellowship
- Ashwin Rajadesingan: Facebook Fellowship
- Sky Wang: NSF GRFP, U-M CSE Outstanding Research Award, CRA Outstanding Undergraduate Researcher, Honorable Mention
- Xingyao Wang: CRA Outstanding Undergraduate Researcher, Honorable Mention
- Carol Zhang: NSF GRFP, CRA Outstanding Undergraduate Researcher, Honorable Mention
- Naitian Zhou: CRA Outstanding Undergraduate Researcher, Honorable Mention

Undergraduate and Masters students

Students appear in rough order of mentorship with the first position after graduation listed, if known.

Jenny Lee (BS CS)
 Kenan Alkiek (MS Information) → UMSI, PhD
 Bohan Zhang (MS CS) → UMSI, PhD
 Jiamin Yang (BS Data Science) → University of Chicago, MS
 Jia Zhu (BS CS)
 Catherine Huang (BS CS)
 Haley Johnson (Bachelors Information)
 Jackson Sargent (BS CS)
 Huiyang Ding (MS Data Science) → Senior Data Engineer (Machine Learning) at Humana
 Lingyun Gao (MS Information)
 Lily Ge (BS CS) → Northwestern University, PhD
 Talia Rizika (BS CogSci)
 Sky Wang (BS CS) → Columbia University, PhD
 Xingyao Wang (BS CS) → University of Illinois, PhD
 Jiajun Bao (BS CS) → Carnegie Mellon University, MS
 Yiming Zhang (BS CS) → University of Chicago (MS)
 Naitian Zhou (BS CS) → University of California, Berkeley, PhD
 Zachary Zipper (BS/MS CS) → Developer at Reservoir Labs
 Justin Chen (BA Sociology + CS) → GaTech, MS
 Lingding Chen (BS CS)
 Rex Chen (BS CS)
 Christopher Demundo (MS Information + MBA)

Shengyu Feng (BS CS) → University of Illinois, MS
 Sayan Ghosh (BS CS) → University of Southern California, PhD
 Thomas Horak (BA Linguistics + CS)
 Sanjana Kolisetty (BS CS)
 Michelle Lee (BS Psychology + CS)
 Carol Zheng (CS) → Carnegie Mellon University, PhD
 Sam Lee (BS CS)
 Trevor Li (BS CS)
 Yaoyang Lin (BS CS)
 Adi Mannari (BS CS)
 Annie Nie (BA CogSci + CS)
 Yiting Shen (BS CS)
 Innocent Ndubuisi-Obi (MS Information) → University of Washington PhD
 Zijian Wang (BS CS) → Stanford University MS

NSF Research Experience for Undergraduates (REU)

2021: Jackson Sergeant (University of Michigan), Athena Aghighi (UC Davis), Michael Geraci (University of Buffalo)
 2020: Sky Wang (University of Michigan)

Girls Encoded Initiative

2021–2022: Taihat Tisha (CS/SI) and Hafsa Sultana (CS)
 2018–2019: Sheily Shah (CS) and Rhea Singh (CS)

Undergraduate Research Opportunity Program (UROP)

2019–2020: Lily Ge, Mahalakshmi Iyer, Martha Miracky, Sanjit Rajendiran

Amazon Alexa Prize Team (2019–2020)

Chung Hoon Hong (MS Data Science), Arushi Jain (MSI), Bruce Park - (MS ECE), Sagnik Sinha Roy (MSI), Yucen Sun (BS CS), Yujian Liu (BS CS), Vihang Agarwal (MS CS), Ryan Draves (BS CS), Zhizhuo Zhou (BS CS + Data Science), Yiting Shen (MS CS), Junjie Xing (MS CS), William Chen (BS CS)

Undergraduate and Masters Theses –

Nikhil Dwibhashyam *Reader*, 2022 (Linguistics, University of Michigan)
 Talia Rizika *Chair*, 2021 (Cognitive Science, University of Michigan)
 Yaoyang Lin *Chair*, 2019 (Computer Science, University of Michigan)
 Yiting Shen *Reader*, 2018 (Computer Science, University of Michigan)

Pre-Faculty Advising

Stanford University – Jade Huang (MS), Lucy Li (SymSys)
McGill University – James McCorriston (BS), Tyler Finnethy (BS), Caitrin Armstrong (BS), Yi-Tian Xu (BS), Hardik Vala (MS), Koustuv Sinha (MS)
Sapienza University – Mohammad Taher Pilehvar (PhD; Co-advised with Roberto Navigli), Moreno Di Vincenzi (MS; Co-advised with Roberto Navigli)
UCLA – Sky Lin, Grace Park, Alex Nau, Moon Bae, Sean August Scot Moon, Alexander Honda, Vasishta Jayanti and David Cohen (all BS)

SERVICE Mentoring

- Rising Stars in CS, 2020
- PhD Recruiting (<https://phd-recruiting.com/>) in Fall 2020. This is an event design for helping students (especially those from non-traditional, underrepresented backgrounds) familiarize themselves with the PhD admissions process and what graduate school entails.
- ACL Mentoring Program 2018, 2019, 2021, 2022

Organizing Committees

Conferences

- Co-Chair for the Midwest Speech and Language Days (MSLD) Symposium 2024.
- Tutorial Chair for ICWSM 2022.
- Sponsorship Chair for ICWSM 2019.
- Data Chair for ICWSM 2017.

Workshops and Shared Tasks

- Workshop on NLP and Computational Social Science (2022) at EMNLP
- Workshop on Teaching NLP (2021) at NAACL
- Workshop on NLP and Computational Social Science (2020) at EMNLP
- Workshop on NLP and Computational Social Science (2019) at NAACL
- Workshop on NLP and Computational Social Science (2017) at ACL
- Workshop on NLP and Computational Social Science (2016) at EMNLP
- Workshop on NLP and Computational Social Science (2016) at ACM WebScience
- Workshop on Semantic Evaluation (SemEval) 2016
- Workshop on Semantic Evaluation (SemEval) 2015
- SemEval-2023 Task 9: Multilingual Tweet Intimacy Analysis
- SemEval-2022 Task 8: Multilingual news article similarity
- SemEval-2016 Task 14: Semantic Taxonomy Enrichment
- SemEval-2014 Task 3: Cross-Level Semantic Similarity
- SemEval-2013 Task 12: Multilingual Word Sense Disambiguation
- SemEval-2013 Task 13: Word Sense Induction for Graded and Non-Graded Senses
- SemEval-2012 Task 2: Measuring Degrees of Relational Similarity

Reviewing

Journals: Science, Nature, PNAS, Artificial Intelligence, Behavior Research Methods, Computational Linguistics, International Journal of Corpus Linguistics, Language Resources and Evaluation, Natural Language Engineering, PLOS ONE, Transactions on Knowledge Discovery from Data (TKDD), Computer Speech & Language

Best Paper Committee ACL (2023), EACL (2024), EMNLP (2024)

Senior Area Chair ACL (2021, 2022, 2024), NAACL (2025)

Area Chair ACL (2018, 2019, 2020), EMNLP (2019, 2021, 2022), EACL (2023), ACL Rolling Review (most of the time)

Senior Program Committee ICWSM (2016, 2017, 2018, 2019), WWW (2020)

Program Committee: ACL (2013, 2016, 2017, 2018), NAACL (2013, 2015, 2016, 2017, 2021, 2023), EMNLP (2015, 2016, 2017), ICWSM (2013–2021), *SEM (2012, 2013, 2016, 2017), SemEval (2013–2016), EACL (2014), LREC (2014, 2016, 2020), WWW (2014–2023), IJCAI (2016), AAAI (2016, 2017, 2018) Workshop on Multilingual and Cross-lingual Methods in NLP (2016), Workshop on Evaluating Vector-Space Representations for NLP (2016), Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages (2016), Workshop on Ethics in NLP (2017), Widening NLP (2018, 2019)

External Reviewer CHI (2016, 2017, 2018), CSCW (2017)

University of Michigan

- GenAI Task Force Project Research Committee Report, College of Engineering (2024)
- University of Michigan, High Performance Computing Committee (HPCC) (2024)

- UMSI / LSA, [Masters of Data Science](#), Committee Member (2022–2023)
- Gerald R. Ford School of Public Policy, Faculty Search Committee (2021–2022)
- CSE, Welcome Committee (2024)
- UMSI, Faculty Search Committee (2021–2022, 2023–2024)
- UMSI, Data Science / Computational Social Science (DS/CSS) seminar co-organizer (Winter 2019)
- UMSI, Dean’s Advisory Committee (2019–2020)
- UMSI, Programming curriculum task force (2018–2019)
- UMSI, Online programs committee (2018–2019)
- UMSI, United Way Representative (2018–2019)
- University of Michigan, Faculty Sanity co-organizer (2018–*present*)