Dr. Aron Culotta – Curriculum Vitae

Associate Professor
Department of Computer Science
Tulane University
New Orleans, LA 70118

http://www.cs.tulane.edu/~aculotta/
aculotta@tulane.edu
Office: ST 303B

Research Focus

Designing statistical language processing algorithms to discover socially-beneficial knowledge from text, particularly social media.
Topics: social network analysis, natural language processing, data mining, machine learning

Contents

Honors/Awards

- Teacher of the Year, 2018, Illinois Institute of Technology, Computer Science Department
- Nayar Prize II Finalist, "Cyberbullying Early Warning and Response System," 2016
- "Outstanding Paper Honorable Mention," AAAI-2015 (1 of 531 accepted papers)
- "Best Paper Honorable Mention," CSCW-2014 (15 of 134 accepted papers)
- "Outstanding Paper Honorable Mention," AAAI-2004 (2 of 121 accepted papers)
- Microsoft Live Labs Fellow, 2006-2008 (full graduate studies tuition plus stipend)

Education

University of Massachusetts at Amherst Amherst, MA
Ph.D., Computer Science, 2008 *advisor: Andrew McCallum*

University of Massachusetts at Amherst Amherst, MA
M.S., Computer Science, 2004

Tulane University New Orleans, LA
B.S., Computer Science (Math minor), *summa cum laude*, 2002

Professional History

Associate Professor of Computer Science September 2020 -
Tulane University New Orleans, LA

Leads the Text Analysis in the Public Interest lab, conducting research in natural language processing, social network analysis, and machine learning.

Associate Professor of Computer Science September 2019 - 2020
Assistant Professor of Computer Science September 2013 - 2019
Illinois Institute of Technology Chicago, IL

Assistant Professor of Computer Science September 2012 - September 2013
Northeastern Illinois University Chicago, IL

Taught three undergraduate courses per semester, including Programming II, Web Programming and Client-side Development.

Assistant Professor of Computer Science January 2009 - May 2012
Southeastern Louisiana University Hammond, LA

Taught three undergraduate courses per semester, including Machine Learning, Theory of Computation, Algorithms, and Simulation.

Chief Scientist January 2007 - present
IT.com Washington, DC

Designed and implemented large-scale statistical topic models for knowledge discovery from email and social media data. Since developing the core intellectual property, current responsibilities are mostly limited to scientific advising.

Software Engineer September 2008 - December 2008
Amazon.com Seattle, WA

Designed product attribute extraction algorithms to enhance the product catalog. Provide scalable, practical algorithms for large, real-world data sets.

Research Assistant September 2002 - May 2008
University of Massachusetts Amherst, MA

Conducted research in machine learning, information extraction, and knowledge discovery in the *Information Extraction and Synthesis Lab* with Professor Andrew McCallum.

Research Intern June 2007 - August 2007
Microsoft Research Redmond, WA

Designed statistical machine learning algorithm to extract and synthesize information from search results.

Research Intern September 2005 - December 2005
Google, Inc. New York, NY

Designed machine learning algorithm to combine relation extraction and knowledge discovery from Wikipedia documents.

Research Intern June 2003 - September 2003
International Business Machines Yorktown Heights, NY

Developed novel support vector machine algorithm to extract relations between people and organizations in newswire text.

Research Assistant June 2001 - September 2001
University of Alabama Huntsville, AL

Optimized memory allocation algorithms for Java's Virtual Machine. Summer program sponsored by National Science Foundation.

Research Funding [NSF Profile]

role	years	amount (to home univ.)	source	title	co-Is
PI	2023–2026	$600,000 ($600,000)	Tulane Office of Research	Tulane Center of Excellence in Community-Engaged Artificial Intelligence	Nick Mattei, Alessandra Bazzano, Caryn Bell, Andrea Boyles, Patrick Button
co-PI	2022–2023	$5,000 ($5,000)	Tulane - Newcomb Institute	Visualizing Raphael's Renaissance Network: A Digital Humanities Collaboration	Alexis Culotta (PI)
PI	2023–2023	$25,000 ($19,990)	NSF-IUCRC	IUCRC Planning Grant: Tulane: Center for Applied Artificial	Nick Mattei, Allan Ding, Matt Montemore, Jihun Hamm, Henry Chu, Zizhan Zheng
PI	2019–2022	$299,995 ($89,995)	NSF-HDBE *funding rate=14%*	Collaborative Research: Predicting Real-time Population Behavior during Hurricanes Synthesizing Data from Transportation Systems and Social Media	Samiul Hasan, Claire Knox, Naveen Eluru U. Central Florida
co-PI	2019–2021	$299,871 ($299,871)	NSF-IIS-EAGER	Understanding the Relationship between Algorithmic Transparency and Filter Bubbles in Online Media	Mustafa Bilgic (PI) Matthew Shapiro IIT
PI	2019-2020	$25,000 ($25,000)	ERIF (IIT)	Social Media Analysis of Indicators of Eating Disorder Treatment Seeking Behavior	Alissa Haedt-Matt IIT (Psych)
PI	2016–2019	$471,992 ($471,992)	NSF-IIS *funding rate=15%*	Quantifying Multifaceted Perception Dynamics in Online Social Networks	Jennifer Cutler Northwestern-Kellogg
PI	2015–2018	$499,251 ($304,725)	NSF-IIS *funding rate=14%*	Reducing Classifier Bias in Social Media Studies of Public Health	Sherry Emery NORC
co-PI	2016-2017	$100,000 ($100,000)	Nayar II	Cyberbullying Early Warning and Response System	Libby Hemphill U. Michigan
PI	2014	$25,000 ($25,000)	ERIF (IIT)	Tracking perception dynamics in online social networks	Jennifer Cutler Northwestern-Kellogg
PI	2010–2013	$109,587 ($109,587)	Louisiana Board of Regents *ranked 1st of 150 submissions*	Discovering Socially Valuable Trends by Extracting Personal Experiences from the Web	
Total:		**$2,435,696 ($2,056,170)**			

Publications

13 journal articles, 41 refereed conference proceedings, 17 refereed workshop proceedings

Google Scholar statistics (1/2023): 6,784 citations h-index=34 i10-index=56
https://scholar.google.com/citations?user=481oUzkAAAAJ
To briefly summarize my publication record, below I list the number of publications in top-tier conferences in each area:

- **Artificial Intelligence:** AAAI (8), IJCAI (2)
- **Web and Social Media Analysis:** ICWSM (5), ASONAM (2), WWW (2)
- **Natural Language Processing:** HLT/NAACL (4), ACL (1), EMNLP (1)
- **Data Mining/Machine Learning:** ICML (1), KDD (1), ICDM (1), SDM (2), CIKM (1)
- **Human-Computer Interaction:** CHI (1), CSCW (1)

Underlined names below indicated student authors.

Thesis

T1 **Aron Culotta**. *Learning and inference in weighted logic with application to natural language processing*. PhD thesis, University of Massachusetts, May 2008. (17 citations in Google Scholar).

Journal Publications

J1 Kamol Chandra Roy, Samiul Hasan, Aron Culotta, and Naveen Eluru. Predicting traffic demand during hurricane evacuation using real-time data from transportation systems and social media. *Transportation Research Part C: Emerging Technologies*, 131:103339, 2021.

J2 Xuan Song, Haoran Zhang, Rajendra A. Akerkar, Huawei Huang, Song Guo, Lei Zhong, Yusheng Ji, Andreas Lothe Opdahl, Hemant Purohit, Andre Supkin, Akshay Pottathil, and **Aron Culotta**. Big data and emergency management: Concepts, methodologies, and applications. *IEEE Transactions on Big Data*, 2020.

J3 Jennifer Cutler and **Aron Culotta**. Using weak supervision to scale the development of machine-learning models for social media-based marketing research. *Applied Marketing Analytics*, 5(2), 2019.

J4 Virgile Landeiro and **Aron Culotta**. Robust text classification under confounding shift. *Journal of Artificial Intelligence Research*, 63, 2018.

J5 Ehsan Mohammady Ardehaly and **Aron Culotta**. Learning from noisy label proportions for classifying online social data. *Social Network Analysis and Mining*, 8(1):2–22, 2018. (h5-index=21).

J6 Jennifer Cutler and **Aron Culotta**. Using online social networks to measure consumers' brand perception. *Applied Marketing Analytics*, 2(4):312–321, 2017.

J7 Libby Hemphill, **Aron Culotta**, and Matthew Heston. #Polar scores: Measuring partisanship using social media content. *Journal of Information Technology & Politics*, 1(1):1–13, 2016. (h-index=24).

J8 **Aron Culotta**. Training a text classifier with a single word using Twitter lists and domain adaptation. *Social Network Analysis and Mining*, 6(1):1–15, 2016. (h5-index=21).

J9 **Aron Culotta**, Nirmal Kumar Ravi, and Jennifer Cutler. Predicting Twitter user demographics using distant supervision from website traffic data. *Journal of Artificial Intelligence Research*, 55:389–408, 2016. (3.371 impact factor; 27 citations in Google Scholar).

J10 **Aron Culotta** and Jennifer Cutler. Mining brand perceptions from Twitter social networks. *Marketing Science*, 2016. (2.36 impact factor; 111 citations in Google Scholar).

J11 **Aron Culotta**, Jennifer Cutler, and Junzhe Zheng. Finding truth in cause-related advertising: A lexical analysis of brands' health, environment, and social justice communications on Twitter. *The Journal of Values-Based Leadership*, 8(2), 2015.

J12 **Aron Culotta**. Lightweight methods to estimate influenza rates and alcohol sales volume from Twitter messages. *Language Resources and Evaluation, Special Issue on Analysis of Short Texts on the Web*, 2013. (.619 impact factor; 81 citations in Google Scholar).

J13 **Aron Culotta**, Trausti Kristjansson, Andrew McCallum, and Paul Viola. Corrective feedback and persistent learning for information extraction. *Artificial Intelligence*, 170:1101–1122, 2006. (3.371 impact factor; 74 citations in Google Scholar).

Refereed Conference Publications

C1 Linsen Li, **Aron Culotta**, Douglas N. Harris, and Nicholas Mattei. Online reviews are leading indicators of changes in k-12 school attributes. In *Proceedings of the ACM Web Conference*, 2023. (365/1900=19.2% accepted).

C2 Karthik Shivaram, Ping Liu, Matthew Shapiro, Mustafa Bilgic, and **Aron Culotta**. Reducing cross-topic political homogenization in content-based news recommendation. In *Proceedings of the 16th ACM Conference on Recommender Systems*, 2022. (39/231=26.9% accepted).

C3 Xintian Li, Samiul Hasan, and **Aron Culotta**. Identifying hurricane evacuation intent on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 618–627, 2022.

C4 Siva K Balasubramanian, Mustafa Bilgic, **Aron Culotta**, Libby Hemphill, Libby, Anita Nikolich, and Matthew A Shapiro. Leaders or followers? a temporal analysis of tweets from ira trolls. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 2–11, 2022.

C5 Wang, Zhao, Kai Shu, and Aron Culotta. Enhancing model robustness and fairness with causality: A regularization approach. In *EMNLP First Workshop on Causal Inference & NLP*, 2021.

C6 Ping Liu, Karthik Shivaram, Matthew Shapiro, **Aron Culotta**, and Mustafa Bilgic. The interaction between political typology and filter bubbles in news recommendation algorithms. In *Proceedings of the Web Conference 2021*, 2021. (357/1736=20.6% accepted).

C7 Zhao Wang and **Aron Culotta**. Robustness to spurious correlations in text classification via automatically generated counterfactuals. In *Proceedings of the Thirty Fifth National Conference on Artificial Intelligence (AAAI 2021*, 2021. (1,692/7,911=21% accepted).

C8 Zhao Wang and **Aron Culotta**. Identifying spurious correlations for robust text classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020.

C9 Bahar Radfar, Karthik Shivaram, and **Aron Culotta**. Characterizing variation in toxic language by social context. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 959–963, 2020.

C10 Zhao Wang and **Aron Culotta**. When do words matter? Understanding the impact of lexical choice on audience perception using individual treatment effect estimation. In *Proceedings of the Thirty Third National Conference on Artificial Intelligence (AAAI 2019)*, 2019. (1150/7095=16% accepted).

C11 Virgile Landeiro, Tuan Tran, and **Aron Culotta**. Discovering and controlling for latent confounds in text classification using adversarial domain adaptation. In *Proceedings of the SIAM International Conference on Data Mining (SDM19)*, 2019. (90/397=22.7% accepted).

C12 Virgile Landeiro and **Aron Culotta**. Collecting representative samples from a search engine by adaptive query generation. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2019.

C13 Tung Nguyen, Li Zhang, and **Aron Culotta**. Estimating tie strength in follower networks to measure brand perceptions. In *Proceedings of the 2019 International Symposium on Foundations and Applications of Big Data Analytics (ASONAM/FAB)*, 2019.

C14 Zhao Wang, Anna Sapienza, **Aron Culotta**, and Emilio Ferrara. Personality and behavior in role-based online games. In *Proceedings of the 2019 IEEE Conference on Games (COG)*, 2019.

C15 Ping Liu, Joshua Guberman, Libby Hemphill, and **Aron Culotta**. Forecasting the presence and intensity of hostility on instagram using linguistic and social features. In *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)*, 2018. (48/295=16% accepted).

C16 Ehsan Ardehaly and **Aron Culotta**. Mining the demographics of political sentiment from Twitter using learning from label proportions. In *Proceedings of the 17th IEEE International Conference on Data Mining (ICDM)*, 2017. (155/778=19.9% accepted).

C17 Virgile Landeiro and **Aron Culotta**. Controlling for unobserved confounds in classification using correlational constraints. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)*, 2017.

C18 Shreesh Kumara Bhat and **Aron Culotta**. Identifying leading indicators of product recalls from online reviews using positive unlabeled learning and domain adaptation. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)*, 2017.

C19 Ehsan Mohammady Ardehaly and **Aron Culotta**. Cold-start recommendations for audio news stories using matrix factorization. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2016. (573/2294=25% accepted).

C20 Ehsan Mohammady Ardehaly and **Aron Culotta**. Domain adaptation for learning from label proportions using self-training. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2016. (573/2294=25% accepted).

C21 Virgile Landeiro and **Aron Culotta**. Robust text classification in the presence of confounding bias. In *Thirtieth National Conference on Artificial Intelligence (AAAI)*, 2016. (549/2132=26% accepted; 28 citations in Google Scholar).

C22 Ehsan Mohammady Ardehaly and **Aron Culotta**. Inferring latent attributes of Twitter users with label regularization. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL)*, 2015. (117/402=29% accepted; 11 citations in Google Scholar).

C23 Virgile Landeiro Dos Reis and **Aron Culotta**. Using matched samples to estimate the effects of exercise on mental health from Twitter. In *Twenty-ninth National Conference on Artificial Intelligence (AAAI)*, 2015. (531/1991=27% accepted; 42 citations in Google Scholar).

C24 **Aron Culotta**, Nirmal Ravi Kumar, and Jennifer Cutler. Predicting the demographics of Twitter users from website traffic data. In *Twenty-ninth National Conference on Artificial Intelligence (AAAI)*, 2015. (531/1991=27% accepted, **Outstanding Paper Honorable Mention** (given to 1 of 531 accepted papers); 118 citations in Google Scholar).

C25 **Aron Culotta**. Reducing sampling bias in social media data for county health inference. In *JSM Proceedings*, 2014. (12 citations in Google Scholar).

C26 Maria E Ramirez-Loaiza, **Aron Culotta**, and Mustafa Bilgic. Anytime active learning. In *Twenty-eigth National Conference on Artificial Intelligence (AAAI)*, 2014. (398/1406=28% accepted; 8 citations in Google Scholar).

C27 Zahra Ashktorab, Christopher Brown, Manojit Nandi, and **Aron Culotta**. Tweedr: Mining Twitter to inform disaster response. In *the 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*, 2014. (46% accepted; 157 citations in Google Scholar).

C28 **Aron Culotta**. Estimating county health statistics with Twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2014. (23% accepted; 98 citations in Google Scholar).

C29 Reid Priedhorsky, **Aron Culotta**, and Sara Y. Del Valle. Inferring the origin location of tweets with quantitative confidence. In *17th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*, 2014. (134/497=27% accepted; **Best Paper Honorable Mention** (given to 15 of 134 accepted papers); 73 citations in Google Scholar).

C30 Michael Wick, Khashayar Rohanimanesh, Kedar Bellare, **Aron Culotta**, and Andrew McCallum. Samplerank: Training factor graphs with atomic gradients. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2011. (152/589=26% accepted; 50 citations in Google Scholar).

C31 Michael Wick, **Aron Culotta**, Khashayar Rohanimanesh, and Andrew McCallum. An entity-based model for coreference resolution. In *SIAM International Conference on Data Mining (SDM)*, 2009. (55/351=16% accepted; 47 citations in Google Scholar).

C32 **Aron Culotta**, Michael Wick, Robert Hall, Matthew Marzilli, and Andrew McCallum. Canonicalization of database records using adaptive similarity measures. In *Proceedings of the*

13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Jose, CA, 2007. (92/513=18% accepted; 20 citations in Google Scholar).

C33 **Aron Culotta**, Michael Wick, Robert Hall, and Andrew McCallum. First-order probabilistic models for coreference resolution. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL)*, pages 81–88, 2007. (72/298=24% accepted; **196 citations** in Google Scholar).

C34 Michael Wick, **Aron Culotta**, and Andrew McCallum. Learning field compatibilities to extract database records from unstructured text. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 603–611, Sydney, Australia, 2006. (18% accepted; 30 citations in Google Scholar).

C35 **Aron Culotta**, Andrew McCallum, and Jonathan Betz. Integrating probabilistic extraction models and data mining to discover relations and patterns in text. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL)*, pages 296–303, New York, NY, June 2006. (62/257=24% accepted; **203 citations** in Google Scholar).

C36 **Aron Culotta** and Andrew McCallum. Joint deduplication of multiple record types in relational data. In *2005 ACM International Conference on Information and Knowledge Management (CIKM)*, pages 257–258, 2005. (106/425=25% accepted; 83 citations in Google Scholar).

C37 **Aron Culotta** and Andrew McCallum. Reducing labeling effort for structured prediction tasks. In *The Twentieth National Conference on Artificial Intelligence (AAAI)*, pages 746–751, Pittsburgh, PA, 2005. (148/803=18% accepted for oral presentation; **227 citations** in Google Scholar).

C38 **Aron Culotta** and Jeffery Sorensen. Dependency tree kernels for relation extraction. In *42nd Annual Meeting of the Association for Computational Linguistics (ACL)*, Barcelona, Spain, 2004. (88/348=25% accepted; **884 citations** in Google Scholar).

C39 Trausti Kristjannson, **Aron Culotta**, Paul Viola, and Andrew McCallum. Interactive information extraction with constrained conditional random fields. In *Nineteenth National Conference on Artificial Intelligence (AAAI)*, San Jose, CA, 2004. (121/453=26% accepted, **Outstanding Paper Honorable Mention** (given to 2 of 121 accepted papers); **186 citations** in Google Scholar).

C40 **Aron Culotta** and Andrew McCallum. Confidence estimation for information extraction. In *Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL)*, Boston, MA, 2004. (43/168=26% accepted; **135 citations** in Google Scholar).

C41 **Aron Culotta**, Ron Bekkerman, and Andrew McCallum. Extracting social networks and contact information from email and the web. In *First Conference on Email and Anti-Spam (CEAS)*, Mountain View, CA, 2004. (35% accepted; **368 citations** in Google Scholar).

Refereed Workshop Publications

W1 **Aron Culotta**, Ginger Zhe Jin, Yidan Sun, and Liad Wagman. Safety reviews on airbnb: An information tale. In *The Platform Strategy Research Symposium*, 2022.

W2 Wang, Zhao, Kai Shu, and Aron Culotta. Enhancing model robustness and fairness with causality: A regularization approach. In *EMNLP First Workshop on Causal Inference & NLP*, 2021.

W3 Ehsan Ardehaly and **Aron Culotta**. Co-training for demographic classification using deep learning from label proportions. In *Proceedings of the ACUMEN Workshop at the 17th IEEE International Conference on Data Mining (ICDM)*, 2017.

W4 Zhao Wang, Jennifer Cutler, and **Aron Culotta**. Are words commensurate with actions? Quantifying commitment to a cause from online public messaging. In *Proceedings of the ACUMEN Workshop at the 17th IEEE International Conference on Data Mining (ICDM)*, 2017.

W5 **Aron Culotta**. Towards identifying leading indicators of smoking cessation attempts from social media. In *Workshop on Computational Health Science at the IEEE International Conference on Healthcare Informatics*, 2016.

W6 Virgile Landeiro and **Aron Culotta**. Reducing confounding bias in observational studies that use text classification. In *AAAI Spring Symposium on Observational Studies through Social Media and Other Human-Generated Content*, 2016.

W7 Elaine Cristina Resende and **Aron Culotta**. A demographic and sentiment analysis of e-cigarette messages on Twitter. In *Workshop on Computational Health Science at the 6th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2015.

W8 Ehsan Mohammady and **Aron Culotta**. Using county demographics to infer attributes of Twitter users. In *ACL Joint Workshop on Social Dynamics and Personal Attributes in Social Media*, 2014. (22 citations in Google Scholar).

W9 Maria E. Ramirez-Loaiza, **Aron Culotta**, and Mustafa Bilgic. Towards anytime active learning: Interrupting experts to reduce annotation costs. In *KDD Workshop on Interactive Data Exploration and Analytics (IDEA)*, 2013. (7 citations in Google Scholar).

W10 Francisco Iacobelli and **Aron Culotta**. Too neurotic, not too friendly: Structured personality classification on textual data. In *ICWSM Workshop on Personality Classification*, 2013. (12 citations in Google Scholar).

W11 Benjamin Mandel, **Aron Culotta**, John Boulahanis, Danielle Stark, Bonnie Lewis, and Jeremy Rodrigue. A demographic analysis of online sentiment during Hurricane Irene. In *NAACL-HLT Workshop on Language in Social Media*, 2012. (122 citations in Google Scholar).

W12 **Aron Culotta**. Towards detecting influenza epidemics by analyzing Twitter messages. In *KDD Workshop on Social Media Analytics*, 2010. (**605 citations** in Google Scholar).

W13 Michael Wick, Khashayar Rohanimanesh, **Aron Culotta**, and Andrew McCallum. Samplerank: Learning preferences from atomic gradients. In *Neural Information Processing Systems (NIPS) Workshop on Advances in Ranking*, 2009. (31 citations in Google Scholar).

W14 **Aron Culotta**, Pallika Kanani, Robert Hall, Michael Wick, and Andrew McCallum. Author disambiguation using error-driven machine learning with a ranking loss function. In *Sixth International Workshop on Information Integration on the Web (IIWeb-07)*, Vancouver, Canada, 2007. (84 citations in Google Scholar).

W15 **Aron Culotta** and Andrew McCallum. Tractable learning and inference with high-order representations. In *International Conference on Machine Learning Workshop on Open Problems in Statistical Relational Learning*, Pittsburgh, PA, 2006. (14 citations in Google Scholar).

W16 **Aron Culotta** and Andrew McCallum. Practical markov logic containing first-order quantifiers with application to identity uncertainty. In *Human Language Technology Workshop on Computationally Hard Problems and Joint Inference in Speech and Language Processing (HLT/NAACL)*, June 2006. (10 citations in Google Scholar).

W17 **Aron Culotta** and Andrew McCallum. Learning clusterwise similarity with first-order features. In *Neural Information Processing Systems (NIPS) Workshop on the Theoretical Foundations of Clustering*, Whistler, B.C., December 2005.

Unrefereed Workshop Publications

U1 **Aron Culotta**, Andrew McCallum, Bart Selman, and Ashish Sabharwal. Sparse message passing algorithms for weighted maximum satisfiability. In *New England Student Colloquium on Artificial Intelligence (NESCAI)*, Ithaca, NY, 2007.

Technical Reports

TR1 Libby Hemphill, Aron Culotta, and Matthew Heston. Framing in social media: How the US congress uses twitter hashtags to frame political issues. Technical report, 2013.

TR2 **Aron Culotta**. Detecting influenza epidemics by analyzing Twitter messages. Technical report, July 2010. (88 citations in Google Scholar).

TR3 **Aron Culotta** and Andrew McCallum. A conditional model of deduplication for multi-type relational data. Technical Report IR-443, University of Massachusetts, September 2005. (13 citations in Google Scholar).

TR4 **Aron Culotta**, David Kulp, and Andrew McCallum. Gene prediction with conditional random fields. Technical Report UM-CS-2005-028, University of Massachusetts, Amherst, April 2005. (55 citations in Google Scholar).

TR5 **Aron Culotta**. Maximizing cascades in social networks. Technical report, University of Massachusetts, 2003. (9 citations in Google Scholar).

Dissemination

Popular Press Mentions of Research

- "Decoding our chatter", Robert Lee Hotz, *Wall Street Journal*. October 1, 2011.
- "Twitter and Disease Control: The Limits of Algorithmic Prediction". Jared Keller, *TheAtlantic.com*. October 26, 2010.

- "Twitter as Medium and Message", Neil Savage, *Communications of the ACM*. Vol. 54 No. 3, Pages 18–20. 2011.
- "Tracking the flu by tracking tweets", Elizabeth Armstrong Moore, *CNET News*. September 29, 2010.

Invited Talks

- "Learning about society by mining the web," UIC 2019 CRIM Symposium, 3/20/2019
- "Observational studies over social media with machine learning," Toyota Technology Institute Colloquium, Chicago, 5/22/2017
- "Text classification in the wild," University of Chicago Training Program in Applied Analytics, 3/24/2017
- **Tutorial:** "Mining Personal Traits in Social Media," SIAM International Conference of Data Mining, Miami, FL, 5/5/2016
- "Towards classifier-driven observational studies from social media," Univ. of Chicago, Harris School of Public Policy, 4/27/2016
- "Confounding bias in text classification," AAAI Spring Symposium on Observational Studies through Social Media, 3/21/2016, Stanford University
- "Dealing with confounding variables in web-based health studies," Loyola University, Big Boulder Workshop on Using Social Data for Social Good, 11/6/2015
- "Identifying and controlling for confounders in social media analysis," Northwestern University, Computational Social Science Summit, 5/17/2015
- "Towards web-scale observational studies of health," Johns Hopkins University, Center for Language and Speech Processing, Baltimore, MD, 3/6/2015
- "Investigating public health using Twitter," University of Illinois, Institute for Health Research and Policy, Chicago, IL, 11/4/2014
- "Understanding public health using Twitter," DePaul University, Chicago, IL, 9/12/2014
- "Addressing selection bias in social media for estimating county health statistics," Joint Statistical Meetings, Boston, MA, 8/4/2014
- "Health Informatics and Social Media", Environmental Protection Agency, Cincinnati, OH, 9/23/2013
- "Health Informatics and Disaster Planning using Social Media Analysis", Los Alamos National Labs, Los Alamos, NM, 7/6/2012
- "Health Informatics and Disaster Planning using Social Media Analysis", **Keynote Speaker**, International Field Directors and Technologies Conference (IFD&TC), Orlando, FL, 5/22/2012
- "Health Informatics and Disaster Planning using Social Media Analysis", National Opinion Research Center at the University of Chicago, Chicago, IL 4/9/2012
- "Health Informatics and Disaster Planning using Social Media Analysis", Tulane University Computer Science Seminar Series, New Orleans, LA 3/2/2012

Teaching Experience

Tulane University (2020-):

- CMPS 2200 Introduction to Algorithms [link]
- CMPS 3140/6140 Introduction to Artificial Intelligence
- CMPS 3160/6160 Introduction to Data Science
- CMPS 4010/4020 Capstone Project I & II
- CMPS 4620/6620 Artificial Intelligence
- CMPS 4730/6730 Natural Language Processing [link]
- CMPS 7980 Natural Language Processing Independent Study (for Public Health PhD Student)

- Along with Ram Mettu, complete redesign of CMPS 2200 to emphasize parallel algorithms and span analysis. Also introduced numerous technologies, including repl.it and GitHub classroom for pair programming.
- Significant redesign of CMPS 4730/6730 to focus on modern neural network approaches to NLP.
- Developed a Service Learning component to CMPS 4010/4020 to collaborate with local non-profits on data science / AI projects.
- Developed online version of CMPS 4620/6620 for Online MSCS.

Illinois Institute of Technology (2013-2020):

- CS429 Information Retrieval [link] — Spring 2014-2016
- CS579 Online Social Network Analysis [link] — Fall 2014-2019 Spring 2018
- CS585 Natural Language Processing [link] — Spring 2017
- CS595 Machine Learning and Social Media [link] — Fall 2013

I introduced and designed CS579 (and the CS595 that preceded it). For CS429 and CS585, I performed a significant redesign of the course.

Student evaluations

Term	Course	Enrolled	Responses	Instructor	Course
Spring 2021	CMPS4730	12	5	4.6	4.6
Fall 2021	CMPS4620	12	7	4.86	5
Spring 2022	CMPS3140	33	29	4.66	4.34
Fall 2022	CMPS3160	36	31	4.77	4.61

Student Research Advising

PhD Students

Graduated:

- Ehsan Mohammady Ardehaly, 2013-2017, "Lightly supervised machine learning for classifying online social data"
- Virgile Landeiro Dos Reis, 2014-2018 , "Removing confounding bias in text classification"
- Zhao Wang, 2016-2021 , "Language, Perception, and Causal Inference in Online Communication"

Current:

- Karthik Shivaram, 2018- , "Filter Bubbles and Algorithmic Personalization for News Recommendation," ABD
- Xintian Li, 2018- , "Geographical Adaptation of Language Models in Social Media"
- Linsen Li, 2022- (co-advising with Nick Mattei)

Co-Advising / Mentorship:

- Ping Liu "Cyberbullying forecasting," 2016-2017 (one year collaboration for Nayar Prize)
- Lan Wei, 2018-2019 , "Measuring perception in online social networks"
- Simone Skeen (Tulane Public Health), 2023, "Understanding mental health and Long-Covid from online data"

Master's Students (Theses and Projects)

- Shreesh Bhat, "Forecasting product recalls from reviews" *github.com/tapilab/is-shreeshbhat*
- Rojin Babayan, "Studying immigration through Twitter" github.com/tapilab/is-rbabayan
- Michael Drews, "Sports Summarization with Natural Language Generation" *github.com/tapilab/is-mdrews93*
- Olivier Dutfoy, "Fantasy sports forecasting" *github.com/tapilab/is-odutfoy*
- Chihung Hsieh, "Airport wait time estimation using social media", *github.com/tapilab/is-ChesterHsieh*
- Ai Jiang, "Estimating effects of health ads on smoking" *github.com/tapilab/is-ajiang3*
- Mayuri Kadam, "Detecting false health claims online" *github.com/tapilab/is-mayurilk*
- Chandra Kumar, "Image analysis for cyberbullying" *github.com/tapilab/is-mychandru024*
- Silambarasan Madanakumar, "Brand similarity on Twitter" *github.com/tapilab/is-SilambarasanM*
- Karthik Mani, "Text summarization with Deep learning" *github.com/tapilab/is-karthikbmk*
- Harsh Parikh, "Inferring demographics from images" *github.com/tapilab/is-hparik11*
- Lola Priego, "Financial Prediction from Twitter" *github.com/lolapriego/financial_predictor*
- Thomas Theissier, "Interactive labeling of tweets for classification" *github.com/tapilab/theissier*
- Bahar Radfar, "Modeling tie strength and online aggression" *github.com/tapilab/is-mellodi*
- Nirmal Kumar Ravi, "Inferring user demographics from Twitter" *github.com/tapilab/aaai-2015-demographics*
- Than Nguyen, "Removing demographic bias in medical diagnosis systems" *github.com/tapilab/is-Thanh*
- Victor Saint Guilhem, "Tracking French politics on Twitter" *github.com/tapilab/is-hjurin*
- Carol Schmitz, "Detecting verbal violence online" *github.com/tapilab/is-carolgrrr*
- Cyril Trosset, "Estimating Twitter Demographics" *github.com/tapilab/ctrosset*
- Xinzhou Yan, "Learning from label proportions while preserving privacy" *github.com/FreyYann/is-FreyYann*
- Sahand Zeinali, "Understanding marijuana use from online content" *github.com/tapilab/is-szeinali*
- Junzhe Zheng, "Identifying impostors in social media"

Undergraduate Students

- Caroline Casella, "Visualizing Artistic Networks," Newcomb Institute-funded project
- Jacob Lehner, "Course Recommender System," Senior Capstone
- Jonathan Licht, Marisa Long, Anna Schoeny, Ethan Sollender, "A Data Dashboard for New Orleans Magistrate Court," Service Learning project in partnership with CourtWatch NOLA
- Batu El, "Fairness with respect to age in screening algorithms" (Honors Thesis)

- Tum-Tum Adeleye, "Algorithmic Fairness in Public Health" Tulane Research & Innovation Award
- Daniel Ralph, "Filter Bubbles in Online Networks" Tulane Research & Innovation Award
- Gabriel Darley, Tan Dung Ngo, Abraham Messing, "A Chatbot For Families Helping Families," Service Learning capstone for an education non-profit
- Sarah Fox, Bennett Kahn, Ila Keshishian, Victoria Li, "A Data Dashboard for CourtWatch NOLA," Service Learning capstone for a criminal justice non-profit
- Xiao Huang, "Twitter bot detection" *github.com/tapilab/is-xhuang1994*
- Yiming Guo, "Identifying hyped memes online" *github.com/tapilab/is-prefixlt*
- Tung Nguyen, "Estimating Tie Strength in Follower Networks to Measure Brand Perceptions"
- Filipe Tabosa, "Personality and Music Tastes" *github.com/tapilab/filipe*
- Tuan Tran, "Controlling for Latent Confounds with Adversarial Domain Adaptation"
- Elaine Resende, "Analysis of e-cigarette messages on Twitter" *github.com/tapilab/chs-2015-ecig*
- Emily Warman, "Understanding demographics of e-cigarette usage"

Thesis committees

- **PhD**: Di Ma (G. Agam), Maria Ramirez-Loaiza (M. Bilgic), Caner Komurlu (M. Bilgic, Proposal), Dane Wilburne (S. Petrovic, Applied Math), Virgile Landeiro (A. Culotta, Proposal), Ehsan Ardehaly (A. Culotta), Xi Rao (L. Hemphill, Humanities, Proposal), Andrew Roback (L. Hemphill, Humanities), Daniel Giles (S. Laurent-Muehleisen, Physics, Proposal), Junze Han (X.Y. Li / P.J. Wan)
- **MS**: Matthew Heston (L. Hemphill, Humanities), Mayuri Kadam (A. Culotta), Sahand Zeinali (A. Culotta)
- **BS**: Xiao Huang (A. Culotta), Yiming Guo (A. Culotta)

Service Activities

Professional Service

- **Program co-Chair:** ICWSM 2020
- **Co-Organizer:** (with Nick Mattei, Jihun Hamm, Brian Summa, and others): Gulf Coast AI Social at NeurIPS 2022
- **Action Editor:** Association for Computational Linguistics Rolling Review, 2021-
- **Area Chair and Senior Program Committees:** AAAI (2017-), ICHI (2017-), IJCAI (2019-)
- **Panelist**, National Science Foundation, IIS and CCF
- **Managing Editor**, Journal of Machine Learning Research, 2008-2018
- **Editor**, Journal of Medical Internet Research, Special Issue on Mining Health Reports, 2017-
- **Publications Chair**, Neural Information Processing Systems Conference, 2009-2011
- **Online Proceedings Chair**, Neural Information Processing Systems Conference, 2007-2011
- **Workshop Organizer**, WSDM 2017 Workshop on Mining Online Health Reports
- **Program Committee:** AAAI, ACL, CoNLL, ICML, IJCAI, KDD, NAACL/HLT
- **Reviewing:** NIPS, UAI, AISTATS, IEEE Trans. on Knowledge Engineering, IEEE Trans. on Audio, Speech and Language Processing, IEEE Trans. on Information Systems.

University Service

Tulane University (2020-):

- Director, Jurist Center for Artificial Intelligence (2022-)
- Data Literacy Quality Enhancement Plan Committee (2021-2022)
- Data Hub Implementation Committee (2022-)

- Graduate Studies Committee (2020-)
- CS Faculty Search Chair (2021, 2022)
- P&T Committee (2020-)
- Co-Organizer (with Nick Mattei and Edson Cabalfin), Data x Community x Design (DxCxD) Workshop (2022)
- Faculty Mentor

Illinois Institute of Technology (2013-2020):

- co-Director, B.S. in Artificial Intelligence
- co-Director, Masters in Artificial Intelligence
  - helped develop and launch two new Artificial Intelligence degrees
- AI Ethics Working Group Co-Chair, 2019-present
- Graduate studies committee, 2013-2018
- Undergraduate studies committee, 2019-present
- Seminar and DLS co-coordinator, 2014-present
- Faculty search committee, 2016-2019
- Admissions committee, 2015-2016
- CAMRAS scholarship interviewer, 2018
- Academic advising: In Fall 2017, advisor of 4 PhD, 62 Masters, 13 Undergraduate students.