Dr. Aron Culotta – Curriculum Vitae

Associate Professor
Department of Computer Science
Illinois Institute of Technology
Chicago, IL 60616

http://www.cs.iit.edu/~culotta/
aculotta@iit.edu
Office: SB 209C 312-567-5261

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

- 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)
- Teacher of the Year, 2018, Illinois Institute of Technology, Computer Science Department

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

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

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

role	years	**amount** (to home univ.)	source	title	co-PIs
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:		**$1,830,696 ($1,426,170)**			

Publications

10 journal articles, 32 refereed conference proceedings, 15 refereed workshop proceedings
[Since joining IIT: 9 journal articles, 17 refereed conference proceedings, 8 refereed workshop proceedings]

Google Scholar statistics: 4,743 citations h-index=28 i10-index=43
https://scholar.google.com/citations?user=481oUzkAAAAJ

A note on conference publications: In computer science, publications in high-quality conference proceedings undergo rigorous peer review and are generally considered the most important measure of research impact. As a report from the Computing Research Association notes, "conference publication is preferred to journal publication, and the premier conferences are generally more selective than the premier journals."[1] To briefly summarize my publication record, below I list the number of publications in top-tier conferences in each area:

- **Artificial Intelligence:** AAAI (7), IJCAI (2)
- **Web and Social Media Analysis:** ICWSM (3), ASONAM (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)

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

J2 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).

J3 Jennifer Cutler and **Aron Culotta**. Using online social networks to measure consumers brand perception. *Applied Marketing Analytics*, 2(4):312–321, 2017.

J4 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).

J5 **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).

J6 **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).

[1] Computing Research Association, "Evaluating Computer Scientists and Engineers for Promotion and Tenure", September 1999.

J7 **Aron Culotta** and Jennifer Cutler. Mining brand perceptions from Twitter social networks. *Marketing Science*, 2016. (2.36 impact factor; 22 citations in Google Scholar).

J8 **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.

J9 **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; 56 citations in Google Scholar).

J10 **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 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).

C2 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).

C3 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.

C4 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.

C5 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.

C6 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).

C7 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).

C8 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.

C9 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.

C10 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).

C11 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).

C12 Virgile Landeiro and **Aron Culotta**. Robust text classification in the presence of confounding variables. In *Thirtieth National Conference on Artificial Intelligence (AAAI)*, 2016. (549/2132=26% accepted).

C13 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).

C14 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; 30 citations in Google Scholar).

C15 **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); 66 citations in Google Scholar).

C16 **Aron Culotta**. Reducing sampling bias in social media data for county health inference. In *JSM Proceedings*, 2014. (12 citations in Google Scholar).

C17 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).

C18 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; 58 citations in Google Scholar).

C19 **Aron Culotta**. Estimating county health statistics with Twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2014. (23% accepted; 59 citations in Google Scholar).

C20 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); 51 citations in Google Scholar).

C21 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).

C22 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).

C23 **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).

C24 **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; **176 citations** in Google Scholar).

C25 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).

C26 **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; **156 citations** in Google Scholar).

C27 **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).

C28 **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; **167 citations** in Google Scholar).

C29 **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; **744 citations** in Google Scholar).

C30 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); **174 citations** in Google Scholar).

C31 **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; **118 citations** in Google Scholar).

C32 **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; **311 citations** in Google Scholar).

Refereed Workshop Publications

W1 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.

W2 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.

W3 **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.

W4 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.

W5 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.

W6 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).

W7 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).

W8 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).

W9 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. (73 citations in Google Scholar).

W10 **Aron Culotta**. Towards detecting influenza epidemics by analyzing Twitter messages. In *KDD Workshop on Social Media Analytics*, 2010. (**436 citations** in Google Scholar).

W11 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).

W12 **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).

W13 **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).

W14 **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).

W15 **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 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. (51 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

Illinois Institute of Technology (2013-present):

- 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 (IIT)	Course (IIT)
Fall 2013	CS595-02	18	5	4.60 (4.26)	4.60 (4.16)
Spring 2014	CS429-01/02/03	25	19	4.84 (4.30)	4.47 (4.17)
Fall 2014	CS579-01/02/03	29	22	4.59 (4.28)	4.64 (4.16)
Spring 2015	CS429-01/02/03	34	23	4.65 (4.29)	4.61 (4.19)
Fall 2015	CS579-01	88	62	4.85 (4.27)	4.76 (4.19)
Fall 2015	CS579-02	15	10	4.90 (4.27)	4.90 (4.19)
Spring 2016	CS429-01	69	46	4.70 (4.35)	4.57 (4.24)
Fall 2016	CS579-01	116	53	4.66 (4.33)	4.64 (4.23)
Fall 2016	CS579-02	20	6	5.00 (4.33)	5.00 (4.23)
Fall 2016	CS579-03	8	4	5.00 (4.33)	4.75 (4.23)
Spring 2017	CS585-01	36	26	4.65 (n/a)	4.69 (n/a)
Spring 2017	CS585-02	6	6	5.00 (n/a)	4.67 (n/a)
Fall 2017	CS579-01	101	40	4.72 (n/a)	4.62 (n/a)
Fall 2017	CS579-02	6	4	4.75 (n/a)	4.75 (n/a)
Spring 2018	CS579-01	69	38	4.81 (n/a)	4.86 (n/a)
	total:	**640**	**364**	**wt. avg.: 4.75/5**	**4.68/5**
			new evaluation instrument		
Spring 2019	CS579	115	56	93%	89%

Student Research Advising

Lead advisor for 6 PhD, 22 MS, and 7 undergrad independent research projects since joining IIT.

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"

Current:

- Zhao Wang, 2016- , "Discovering the effect of linguistic style on perception"
- Karthik Shivaram, 2018- , "Understanding social dynamics of cyberbullying"
- Lan Wei, 2018- , "Measuring perception in online social networks"
- Xintian Li, 2018- , "Learning from Label Proportions and adversarial neural networks"

Former:

- Ping Liu "Cyberbullying forecasting," 2016-2017 (one year collaboration for Nayar Prize)

Master's Students (CS597 / CS591 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

- 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 (10)**: 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 (3)**: Matthew Heston (L. Hemphill, Humanities), Mayuri Kadam (A. Culotta), Sahand Zeinali (A. Culotta)
- **BS (2)**: Xiao Huang (A. Culotta), Yiming Guo (A. Culotta)

Service Activities

Professional Service

- **Program Chair:** ICWSM 2020
- **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

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