

Maria Antoniak

✉ antoniakmaria@gmail.com

🌐 <http://maria-antoniak.github.io>

Research Areas

Natural Language Processing • Cultural Analytics
Narratives • Online Communities • Healthcare

Education

- 2016-Present **Ph.D. Information Science**, Cornell University
- Committee: David Mimno, Lillian Lee, Jeff Rzeszotarski, Richard Jean So
 - Minor: Computer Science
- 2013-2014 **M.S. Computational Linguistics**, University of Washington
- Advisor: Fei Xia
 - Thesis: *Extracting Topically Related Synonyms from Twitter*
- 2007-2011 **B.A. Program of Liberal Studies**, University of Notre Dame
- Glynn Family Honors Program

Publications

- 2021 [Bad Seeds: Evaluating Lexical Methods for Bias Measurement](#)
Maria Antoniak and David Mimno
Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)
- 2021 [Tags, Borders, and Catalogs: Social Re-Working of Genre on LibraryThing](#)
Maria Antoniak, Melanie Walsh, and David Mimno
Proceedings of the ACM on Human-Computer Interaction (CSCW)
- 2021 [The Goodreads “Classics”: A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism](#)
Melanie Walsh and **Maria Antoniak**
Journal of Cultural Analytics, Post45 (joint issue)
- 2020 [Imagined Examples of Painful Experiences Provided by Chronic Low Back Pain Patients and Attributed a Pain Numerical Rating Score](#)
Robert Stewart Griffin, **Maria Antoniak**, Phuong Dinh Mac, Vladimir Kramskiy, Seth Waldman, David Mimno
Frontiers in Neuroscience
- 2019 [Narrative Paths and Negotiation of Power in Birth Stories](#)
Maria Antoniak, David Mimno, and Karen Levy
Proceedings of the ACM on Human-Computer Interaction (CSCW)
- 2018 [Evaluating the Stability of Embedding-based Word Similarities](#)
Maria Antoniak and David Mimno
Transactions of the Association for Computational Linguistics (TACL)

- 2016 [Natural Language Processing Techniques on Oil and Gas Drilling Data](#)
Maria Antoniak, Jeff Dagliesh, and Justin Lo
SPE Intelligent Energy Conference and Exhibition
- 2016 [Leveraging Paraphrase Labels to Extract Synonyms from Twitter](#)
Maria Antoniak, Eric Bell, and Fei Xia
Proceedings of the Florida Artificial Intelligence Research Society Conference (FLAIRS)

Workshop Publications

- 2021 ['Tecnologica cosa': Modeling Storyteller Personalities in Boccaccio's 'Decameron'](#)
A. Feder Cooper, **Maria Antoniak**, Christopher De Sa, Marilyn Migiel, and David Mimno
Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL) at Conference on Empirical Methods in Natural Language Processing (EMNLP) (peer reviewed)

Work in Progress

- Under Review* **The Afterlives of Shakespeare and Company in Online Social Readership**
Maria Antoniak, David Mimno, Rosamond Thalken, Melanie Walsh, Matthew Wilkens, Gregory Yauney (all equal contribution)
- Under Review* **Attitudes Towards Self-Disclosure Across Online Gaming Communities: Results From a Large-Scale Survey**
Maria Antoniak, Alexandra Olteanu, Asia Biega, Fernando Diaz, Miriam Meister
- Under Review* **Birth Control Discussions Across Online Platforms: Patterns and Personal Experiences of Side Effects and Pain**
LeAnn McDowall, **Maria Antoniak**, David Mimno
- In Progress* **Twitter's Intimate Publics: #BLM, Affect, Narrative**
Richard Jean So, Long Le-Khac, **Maria Antoniak**

Industry Experience

- 2021 **Twitter**, *Cortex Intern*
- Mentors: Stephen Ragain and Jay Baxter
 - Machine learning to support crowdsourced misinformation identification (Birdwatch)
 - Topic and rating diversity measurements of Birdwatch participants
- 2020-2021 **Microsoft Research**, *FATE Research Intern*
- Mentors: Fernando Diaz, Asia Biega, Alexandra Olteanu
 - Team: Fairness, Accountability, Transparency, Ethics (FATE)
 - Survey (37,000 respondents) about self-disclosure in online gaming communities
 - Automatic measurements of self-disclosures in text

- 2019 **Facebook**, *Core Data Science Intern*
- Mentors: Amit Bahl and Emma Ideal
 - Unsupervised models of conversational roles in online health support and hiring groups
- 2018 **Microsoft Research**, *Research Intern*
- Mentor: Ranjani Ramamurthy
 - Unsupervised extractive summarization of clinical dialogues for Project EmpowerMD
- 2014-2016 **MAANA**, *Data Scientist*
- Data science design and implementation of proof-of-concept projects
 - Product query, named entity recognition, and machine learning pipeline features
- 2014 **Pacific Northwest National Laboratory**, *National Security Intern*
- Mentor: Eric Bell
 - Synonym extraction for paraphrase detection on Twitter
 - Supported other NLP projects (sentiment analysis, topic modeling)
- 2013 **Pacific Northwest National Laboratory**, *Mobile Developer Intern*
- Mentor: Courtney Corley
 - Public health, foodborne illness, mobile application development

Workshop Presentations

- 2022 **Online Readership and Perceptions of Genres Over Time**
 Maria Antoniak, Melanie Walsh, and David Mimno
Digital Humanities (DH)
- 2021 **Bad Seeds: Evaluating Lexical Methods for Bias Measurement**
 Maria Antoniak and David Mimno
 Poster at *Gender Bias Workshop*, collocated with *ACL*
- 2021 **BERT for Humanists: Uses, Challenges, Translations**
 Maria Antoniak, Melanie Walsh, David Mimno, and Matthew Wilkens
Association for Computers and the Humanities (ACH)
- 2019 **Diverging Paths in Birth Stories: A Medical Dataset for Narrative Analysis**
 Maria Antoniak and David Mimno
 Oral presentation at *Workshop for Narrative Understanding, NAACL*
- 2018 **Comparative Analysis of Cultural Biases with Word Embeddings**
 Maria Antoniak and David Mimno
 Poster at *Text as Data (TADA)*
- 2017 **Evaluating the Stability of Embedding-based Word Similarities**
 Maria Antoniak and David Mimno
 Poster at *Amazon Graduate Research Symposium*
- 2015 **Extracting Topically Related Synonyms from Twitter**
 Maria Antoniak, Eric Bell, and Fei Xia
 Poster at *Workshop for Women in Machine Learning (WiML)*, collocated with *NeurIPS*

Invited Talks

- 2022 **Narratives in Online Communities**
Maria Antoniak
Workshop on Narrative Understanding (NAACL)
- 2021 **Modeling Shared Personal Experiences in Online Communities**
Maria Antoniak
MIT, Center for Constructive Communication
- 2021 **Modeling Shared Personal Experiences in Online Communities**
Maria Antoniak
University of California, Berkeley, NLP Seminar
- 2021 **Modeling Shared Personal Experiences in Online Communities**
Maria Antoniak
Stanford University, NLP Seminar
- 2021 **Building Social Bonds in Online Gaming Communities**
Maria Antoniak and Alexandra Olteanu
Microsoft AI & Gaming Research Summit
- 2020 **Narrative Paths and Negotiation of Power in Birth Stories**
Maria Antoniak
University of Michigan, Data Science/Computational Social Science Seminar

Tutorials

- 2022 [BERT for Social Sciences and Humanities](#)
International Conference on the Web and Social Media (ICWSM)
Maria Antoniak, Melanie Walsh, David Mimno, and Matthew Wilkens
- 2021 [Introduction to NLP for Journalists](#)
Brown Institute (Columbia Journalism School and Stanford Engineering)
Maria Antoniak
- 2021 [BERT for Computational Social Science](#)
NLP+CSS 201 Tutorial Series
Maria Antoniak
- 2021 [BERT for Humanists Virtual Workshop](#)
Maria Antoniak, Melanie Walsh, David Mimno, and Matthew Wilkens

Panels

- 2021 **Computational Approaches to Book Reviews**
Matthew Lavin, Melanie Walsh, Maria Antoniak, Dan Sinykin, Jordan Pruett
Association for Computers and the Humanities (ACH)
- 2020 **Replication and Computational Literary Studies**
Christof Schoch, Karina van Dalen-Oskam, Maria Antoniak, Fotis Jannidis, David Mimno
Digital Humanities

- 2020 **Cultural Analytics and the Book Review: Models, Methods, and Corpora**
 Matthew J. Lavin, Kent Chang, Dan Sinykin, Melanie Walsh, Maria Antoniak, Kent Chang,
 Yuerong Hu, Wenyi Shang, Aniruddha Sharma, Shubhangi Singhal, Ted Underwood, Jes-
 sica Witte, and Peizhen Wu
Digital Humanities

Teaching Experience

- Fall 2021 **INFO 2950: Introduction to Data Science**, Cornell University, TA
 Fall 2020 **INFO 3350: Text Mining for History and Literature**, Cornell University, TA
 Fall 2019 **INFO 3350: Text Mining for History and Literature**, Cornell University, TA
 Spring 2019 **CS/INFO 3300: Data-Driven Web Applications**, Cornell University, TA
 Spring 2017 **CS/INFO 3300: Data-Driven Web Applications**, Cornell University, TA
 2011-2012 **Ukrainian Catholic University**, Lviv (Ukraine), English Teacher
 2010 & 2011 **Humanities Spring**, Assisi (Italy), TA
 2010-2011 **First Year Writing Tutorial**, University of Notre Dame, TA
 2008-2011 **University Writing Center**, University of Notre Dame, Writing Tutor

Guest Lectures

- May 2022 **Computational Linguistics I**, Cornell University, Marten van Schijndel
 Sept 2021 **Approaches to Data Science with Text**, Emory University, Lauren Klein
 June 2021 **Computational Text Analysis**, Barnard College, Adam Poliak

Honors, Awards, & Scholarships

- 2022 Graduate Student Excellence in Leadership Award, Honorable Mention, *Diversity Pro-
 grams in Engineering, Cornell University*
 2021 Rising Star in Data Science, *University of Chicago Rising Stars*
 2021 Rising Star in EECS, *MIT EECS Rising Stars*
 2018 Information Science Outstanding Service Award, *Cornell University*
 2016-2017 Information Science Fellowship, *Cornell University*
 2013 1st Place, *PNNL Biosurveillance Mobile App Development Competition*
 2007-2011 Glynn Family Honors Program, *University of Notre Dame*
 2007 National Merit Scholar, *National Merit Scholarship Program*

Professional Service

- 2022 Reviewing for *Digital Humanities (DH)*
 Reviewing for *Student Research Workshop, ACL*
 Reviewing for *FACCT*
 2021 Reviewed for *W-NUT at EMNLP*
 Reviewed for *EMNLP*
 Reviewed for *CSCW*
 Reviewed for *Student Research Workshop, ACL*
 Reviewed for *NAACL*
 Reviewed for *EACL*
 Reviewed for *CHI*

2020 Reviewed for *PLOS ONE*
2020 Reviewed for *EMNLP*

Mentoring

2019-2021 *LeAnn McDowall*, undergraduate; data analysis of birth control websites; won Best Paper Award at the *Black in AI Workshop* at *NeurIPS 2020*.

Community Service

2020-2022 *President*, Graduate Students for Gender Inclusion in Computing
2020-2021 *Lead Organizer*, Computational Social Science with Care Working Group
2020-2021 *Student Representative*, Broadening Participation in Information Science at Cornell
2020-2021 *Webmaster*, Cornell Graduate Women in Science
2019-2020 *President*, Graduate Students for Gender Inclusion in Computing
2019 *Lead Organizer*, Cornell Information Science Research Retreat
2019-2020 *Treasurer*, Cornell Graduate Women in Science
2018-2019 *Social Chair*, Cornell Graduate Women in Science
2017-2019 *Vice President & Treasurer*, Information Science Graduate Student Association
2010-2011 *Editorial Board Member*, Notre Dame Journal of Undergraduate Research

Non-Degree Studies

Summer 2016 Lisbon Machine Learning Summer School (LxMLS), *Lisbon, Portugal*
Summer 2011 Ukrainian Language Summer School, *Lviv, Ukraine*
Spring 2010 French Language & Culture, Université Catholique de l'Ouest, *Angers, France*