

Website: maria-antoniak.github.io

Email: maria.antoniak@colorado.edu

Research Interests: natural language processing, cultural analytics, healthcare

Professional Appointments

- 2025– **Assistant Professor**, University of Colorado Boulder, Computer Science
- 2024–2025 **Postdoc**, University of Copenhagen, Pioneer Centre for AI
- 2022–2024 **Young Investigator**, Allen Institute for AI
- 2023 **Summer Fellow**, ETH Zürich, Department of Humanities, Social, and Political Sciences

Education

- 2016–2022 **PhD in Information Science**, Cornell University
- 2013–2014 **MS in Computational Linguistics**, University of Washington
- 2007–2011 **BA in Program of Liberal Studies**, University of Notre Dame

Publications

- 2025 [so much depends / upon / a whitespace: Why Whitespace Matters for Poets and LLMs](#)
Empirical Methods in Natural Language Processing (EMNLP)
S Bhayravajjula, M Walsh, A Preus, [M Antoniak](#)
- 2025 [Multi-Modal Framing Analysis of News](#)
Empirical Methods in Natural Language Processing (EMNLP)
A Arora, S Yadav, [M Antoniak](#), S Belongie, I Augenstein
- 2025 [LLMs as Research Tools: A Large Scale Survey of Researchers' Usage and Perceptions](#)
Conference on Language Modeling (COLM)
Z Liao, [M Antoniak](#), I Cheong, EY Cheng, A Lee, K Lo, JC Chang, AX Zhang
- 2025 [Research Borderlands: Analysing Writing Across Research Cultures](#)
Association for Computational Linguistics (ACL)
S Bhatt, T August, [M Antoniak](#)
- 2025 [Information-Guided Identification of Training Data Imprint in \(Proprietary\) Large Language Models](#)
North American Chapter of the Association for Computational Linguistics (NAACL)
A Ravichander, J Fisher, T Sorensen, X Lu, Y Lin, [M Antoniak](#), N Miresghallah, C Bhagavatula, Y Choi
- 2024 [Clinician as Editor: Notes in the Era of AI Scribes](#)
The Lancet
S Altschuler, I Huntington, [M Antoniak](#), LF Klein
- 2024 [Sonnet or Not, Bot? Poetry Evaluation for Large Models and Datasets](#)
Empirical Methods in Natural Language Processing (EMNLP) Findings
M Walsh, A Preus, [M Antoniak](#)
- 2024 [The Empirical Variability of Narrative Perceptions of Social Media Texts](#)
Empirical Methods in Natural Language Processing (EMNLP)
J Mire, [M Antoniak](#), E Ash, A Piper, M Sap
- 2024 [NLP for Maternal Healthcare: Perspectives and Guiding Principles in the Age of LLMs](#)
ACM Conference on Fairness, Accountability, and Transparency (FACCT)
[M Antoniak](#), A Naik, CS Alvarado, LL Wang, IY Chen
- 2024 [Sensemaking About Contraceptive Methods Across Online Platforms](#)
International Conference on Web & Social Media (ICWSM)
L McDowall, [M Antoniak](#), D Mimno
- 2024 [Trust No Bot: Personal Disclosures in Human-LLM Conversations](#)
Conference on Language Modeling (COLM)
[M Antoniak](#),* N Miresghallah,* Y More,* Y Choi, G Farnadi (*equal contribution)

- 2024 [Where Do People Tell Stories Online? Story Detection Across Online Communities](#)
Association for Computational Linguistics (ACL)
[M Antoniak](#), J Mire, M Sap, E Ash, A Piper
- 2024 [Personalized Jargon Identification for Enhanced Interdisciplinary Communication](#)
North American Association for Computational Linguistics (NAACL)
Y Guo, JC Chang, [M Antoniak](#), E Bransom, T Cohen, LL Wang, T August
- 2024 [The Afterlives of Shakespeare and Company in Online Social Readership](#)
Journal of Cultural Analytics and Modernism/modernity (joint issue)
[M Antoniak](#), D Mimno, R Thalken, M Walsh, M Wilkens, G Yauney (all equal contributions)
- 2023 [Riveter: Measuring Power and Social Dynamics Between Entities](#)
Association for Computational Linguistics (ACL), Demo Track
[M Antoniak](#), A Field, J Mun, M Walsh, LF Klein, M Sap
- 2023 [#BLM Insurgent Discourse, White Structures of Feeling, and the Fate of the 2020 “Racial Awakening”](#)
Journal of New Literary History
L Le-Khac, [M Antoniak](#), RJ So
- 2021 [Bad Seeds: Evaluating Lexical Methods for Bias Measurement](#)
Association for Computational Linguistics (ACL)
[M Antoniak](#), D Mimno
- 2021 [The Goodreads “Classics”: A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism](#)
Journal of Cultural Analytics, Post45 (joint issue)
M Walsh, [M Antoniak](#)
- 2021 [Tags, Borders, and Catalogs: Social Re-Working of Genre on LibraryThing](#)
ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW)
[M Antoniak](#), M Walsh, D Mimno
- 2020 [Imagined Examples of Painful Experiences Provided by Chronic Low Back Pain Patients and Attributed a Pain Numerical Rating Score](#)
Frontiers in Neuroscience
RS Griffin, [M Antoniak](#), PD Mac, V Kramskiy, S Waldman, D Mimno
- 2019 [Narrative Paths and Negotiation of Power in Birth Stories](#)
ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW)
[M Antoniak](#), K Levy, D Mimno
- 2018 [Evaluating the Stability of Embedding-based Word Similarities](#)
Transactions of the Association for Computational Linguistics (TACL)
[M Antoniak](#), D Mimno

Pre-Prints / Under Review

- 2025 AI Fiction in the Wild
N Gupta, [M Antoniak](#), M Walsh
- 2025 [Epistemic Diversity and Knowledge Collapse in Large Language Models](#)
D Wright, S Masud, J Moore, S Yadav, [M Antoniak](#), PE Christensen, CY Park, I Augenstein
- 2025 [Provocations from the Humanities for Generative AI Research](#)
L Klein, M Martin, A Brock, [M Antoniak](#), M Walsh, JM Johnson, L Tilton, D Mimno
- 2025 [Cultural Evaluations of Vision-Language Models Have a Lot to Learn from Cultural Theory](#)
S Yadav, L Tilton, [M Antoniak](#), T Arnold, J Li, SM Pawar, A Karamolegkou, S Frank, Z An, N Rostamzadeh, D Herscovitch, S Belongie, E Shutova
- 2025 [Meaning Is Not A Metric: Using LLMs to make cultural context legible at scale](#)
C Kommers, D Hemment, [M Antoniak](#), JZ Leibo, H Long, E Robinson, A Sobey

Workshop Publications (Archival)

- 2022 [Narrative Datasets through the Lenses of NLP and HCI](#)
Narrative Understanding Workshop (NAACL)
S Sultana, R Zhang, H Lim, [M Antoniak](#)
- 2021 [Heroes, Villains, and Victims, and GPT-3: Automated Extraction of Character Roles Without Training Data](#)
NLP+HCI Workshop (NAACL)
D Stambach, [M Antoniak](#), E Ash
- 2021 [‘Tecnologica cosa’: Modeling Storyteller Personalities in Boccaccio’s Decameron](#)
Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL) (EMNLP)
AF Cooper, [M Antoniak](#), C De Sa, M Migiel, D Mimno

Grants

- 2026 **Arts and Humanities Research Council (AHRC-UKRI)**, selection for pre-funding workshop (top 7% of applications), final funding announcements in summer 2026
- 2025 **Arts and Humanities Research Council (AHRC-UKRI)**, \$2000 research grant

Honors & Awards

- 2022 Graduate Student Excellence in Leadership Award (Honorable Mention), *Cornell Diversity Programs in Engineering*
- 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 Information Science Fellowship, *Cornell University*
- 2013 1st Place, *PNNL Biosurveillance Mobile App Development Competition*
- 2011 Glynn Family Honors Program, *University of Notre Dame*
- 2007 National Merit Scholar

Academic Service

- 2026 **Publicity Chair**, ACM Conference on Fairness, Accountability, and Transparency (FAccT)
Tutorials Co-Chair, International Conference on Computational Social Science (IC2S2)
Senior Area Chair, Association for Computational Linguistics (ACL)
- 2025 **Advisory Board Member**, Anthology of Computers and the Humanities
Editorial Board Member, Journal of Cultural Analytics
Member, Working Group on AI and Research as part of the MLA Task Force on Generative AI Initiatives
Guest Editor, Computational Humanities Research (CHR) Journal, Special Issue on LLMs
Area Chair, Conference on Language Modeling (COLM)
Ethics Co-Chair, North American Association for Computational Linguistics (NAACL)
Advisory Board Member, Computational Humanities Research (CHR) Journal
Senior Area Chair, Association for Computational Linguistics (ACL)
Publicity Chair, ACM Conference on Fairness, Accountability, and Transparency (FAccT)
Reviewer, TACL, Journal of Cultural Analytics, Workshop on Narrative Understanding, Workshop on NLP for Democracy, Journal of Computational Literary Studies
- 2024 **Advisory Board Member**, Journal of Computational Humanities Research (CHR)
Senior Area Chair, Conference on Computational Linguistics, Language Resources & Evaluation (COLING)
Ethics Co-Chair, North American Association for Computational Linguistics (NAACL)
Workshops Co-Chair, International Conference on Web & Social Media (ICWSM)
Area Editor, ACL Rolling Review (ARR)

Reviewer, COLM, FAccT, JAIR, TACL, Sociological Methods & Research, IC2S2, NLP4DH, Workshop on Narrative Understanding, JMIR, etc.

2023 **Area Chair**, Empirical Methods in Natural Language Processing (EMNLP)

University Service

2025–2026 **Chair**, Committee for Graduate Student Mentorship & Awards

2025–2026 **Member**, Digital Humanities Graduate Certificate Program

Industry Experience

2023 **Association of American Medical Colleges (AAMC)**, Consultant

2021 **Twitter**, Intern, Team: Cortex, Birdwatch (Community Notes)

2020–2021 **Microsoft Research**, Intern, Team: [Fairness, Accountability, Transparency, Ethics \(FATE\)](#)

2019 **Facebook**, Intern, Team: Core Data Science, Healthcare Support Groups

2018 **Microsoft Research**, Intern, Team: [Project EmpowerMD](#)

2014–2016 **Maana**, Data Scientist

2013 **Pacific Northwest National Laboratory**, Intern, Biosurveillance

Tutorials (Designed & Led)

2025 **NLP for Narratives**

COMTEXT 2025

[M Antoniak](#)

2025 **NLP for Narratives**

University of Warsaw, Centre for Excellence in the Social Sciences

[M Antoniak](#)

2023 **Large Language Models for Humanists: A Hands-On Introduction**

University of Washington, Simpson Center for the Humanities

[M Antoniak](#), [M Walsh](#)

2023 **Introduction to NLP: From Story to Data to Action**

Association of American Medical Colleges

[M Antoniak](#)

2023 **Large Language Models for FAccT Researchers**

ACM Conference on Fairness, Accountability, and Transparency (FAccT)

[M Antoniak](#), [M Walsh](#), [L Soldaini](#), [M Wilkins](#), [D Mimno](#)

2023 **NLP Overview**

Intelligence, Data, Ethics and Society (IDEAS) Summer Program (Northeastern University)

[M Antoniak](#)

2023 **Transformers in NLP**

Hertie School, Data Science Summer School

[M Antoniak](#)

2022 **BERT for Social Scientists and Humanists**

Bell Labs, Cambridge

[M Antoniak](#), [M Walsh](#)

2022 **BERT for Social Scientists and Humanists**

International Conference on the Web and Social Media (ICWSM)

[M Antoniak](#), [M Walsh](#), [D Mimno](#), [M Wilkins](#)

2021 **Introduction to NLP for Journalists**

Brown Institute (Columbia Journalism School and Stanford Engineering)

[M Antoniak](#)

2021 **BERT for Computational Social Science**

NLP+CSS 201 Tutorial Series

M Antoniak

2021 **BERT for Computational Social Science**

BERT for Humanists Virtual Workshop

M Antoniak, M Walsh, D Mimno, M Wilkens

Invited Talks

2025 **University of Copenhagen**, Seminar on Digital Literary History

2025 **Santa Fe Institute**, Workshop on Science of Storytelling

2025 **Colgate University**, Mind, Brain, and Behavior Initiative

2025 **Sciences Po**, médialab, conference on *Social Science and Generative AI: Inquiries, Instruments, Consequences*

2025 **Alan Turing Institute**, conference on *Doing AI Differently*

2025 **Linköping University**, Institute for Analytical Sociology

2025 **University of Saarland**, Interdisciplinary Institute for Societal Computing

2024 **University of Stuttgart**, Institute for Natural Language Processing

2024 **Uppsala University**, Statistics Seminar

2024 **Aarhus University**, Interacting Minds Center

2024 **NAACL**, Workshop on NLP+CSS (Keynote)

2024 **Bocconi University**, MilaNLP Seminar Series

2024 **UC San Diego**, DISC and IN STEP Children's Mental Health Research Center

2024 **LREC-COLING**, Workshop on Reference, Framing, and Perspective (Keynote)

2024 **Max Planck Institute**, Institute for Software Systems (MPI SWS)

2024 **University of British Columbia**, NLP Group

2024 **Emory University**, Department of Computer Science

2023 **NLP4DH**, (Keynote)

2023 **ETH Zürich**, Online Seminar in Economics and Data Science

2023 **UMass Amherst**, Machine Learning and Friends Lunch

2023 **Princeton University**, Quantitative Social Science Colloquium (QSSC)

2023 **Educational Testing Service (ETS)**, AI Labs

2023 **American Association of Medical Colleges**, Maternal Health Equity Workshop

2023 **Dartmouth College**, Computational Formalism Workshop

2022 **McGill University**, Narrative and Society Conference

2022 **NAACL**, Workshop on Narrative Understanding (Keynote)

2021 **MIT**, Center for Constructive Communication

2021 **University of California, Berkeley**, NLP Seminar

2021 **Stanford University**, NLP Seminar

2021 **Microsoft**, AI and Gaming Research Summit

2020 **University of Michigan**, Data Science/Computational Social Science Seminar

Panels (Invited)

2023 **Reception History in Many Dimensions: New Research on Book Reviews**

ADHO Digital Humanities Conference

M Antoniak, M Lavin, M Walsh, Y Hu, T Underwood, D Bishop, L Senatorova, W Shang

2021 **Computational Approaches to Book Reviews**

Association for Computers and the Humanities (ACH)

M Antoniak, M Lavin, M Walsh, D Sinykin, J Pruett

2020 **Replication and Computational Literary Studies**

ADHO Digital Humanities Conference

M Antoniak, C Schoch, K van Dalen-Oskam, F Jannidis, D Mimno

2020 **Cultural Analytics and the Book Review: Models, Methods, and Corpora**

ADHO Digital Humanities Conference

M Antoniak, MJ Lavin, K Chang, D Sinykin, M Walsh, Y Hu, W Shang, A Sharma, S Singhal, T Underwood, J Witte, P Wu

Teaching Experience

2025 **NLP for Cultural Analytics**

University of Colorado Boulder, Computer Science Department

2023 **NLP for Cultural Analytics**

University of Washington, Department of Linguistics (Lecturer)

2021 **INFO 2950: Introduction to Data Science**

Cornell University, Department of Information Science (TA)

2020 **INFO 3350: Text Mining for History and Literature**

Cornell University, Department of Information Science (TA)

2019 **INFO 3350: Text Mining for History and Literature**

Cornell University, Department of Information Science (TA)

2019 **CS/INFO 3300: Data-Driven Web Applications**

Cornell University, Department of Information Science (TA)

2017 **CS/INFO 3300: Data-Driven Web Applications**

Cornell University, Department of Information Science (TA)

2011–2012 **English Language & Culture**

Ukrainian Catholic University, Lviv (Lecturer)

2010 **Poetry, Creative Writing**

Humanities Spring, Assisi (TA)

2008–2011 **Writing Tutoring**

University of Notre Dame, University Writing Center (Tutor)

Guest Lectures

2022 **Computational Linguistics I**

Cornell University, Marten van Schijndel

2021 **Approaches to Data Science with Text**

Emory University, Lauren Klein

2021 **Computational Text Analysis**

Barnard College, Adam Poliak

Workshop Attendance (Invited)

2025 *Doing AI Differently*, hosted by the Alan Turing Institute in London

2025 *AI STORIES*, hosted by the University of Bergen in Solstrand

2025 *Impressed by Reading: Measuring the Impact*, hosted by Leiden University, Lorentz Center in Leiden

2024 *Humanities and AI Virtual Institute*, hosted by Schmidt Sciences at Oxford University

2019 *NovelTM: Text Mining the Novel*, hosted by McGill University in Banff

Professional Trainings

2023 Narrative Medicine Workshop, Columbia University

2016 Lisbon Machine Learning Summer School (LxMLS)

Mentoring

- 2024– Teagan Johnson, PhD student at CU Boulder (primary advisor)
- 2024–2025 [Sriharsh Bhyravajjula](#), project on whitespace usage in poetry
- 2024–2025 [Uma Gunturi](#), project on counter narratives in social media conversations
- 2024–2025 [Shaily Bhatt](#), intern at AI2; PhD student at CMU working with Fernando Diaz; project on evaluating models for adaptations to research communities and cultures
- 2023–2024 [Hang Jiang](#), intern at AI2; PhD student at MIT working with Deb Roy; project on automatic measurement of research values in computer science papers
- 2023–2024 [Yujia Gao](#); project on measuring values expressed in Goodreads book reviews; accepted to PhD program at Brown University
- 2019–2024 [LeAnn McDowall](#), undergraduate at Cornell University; data analysis of contraception discussions; won Best Paper Award at the *Black in AI Workshop* at *NeurIPS* (2020); awarded the NSF *CSGrad4US* Fellowship (2022); paper published at ICWSM

Selected Community Service

- 2020–2022 **Graduate Students for Gender Inclusion in Computing**, President & Founder
Cornell University
- 2020–2021 **Computational Social Science with Care Working Group**, President & Lead Organizer
- 2020–2021 **Broadening Participation Committee**, Student Representative
Cornell University, Department of Information Science
- 2020–2021 **Graduate Women in Science**, Webmaster, Retreat Founder
Cornell University