

Katherine Atwell

PhD Candidate, Northeastern University

✉ atwell.ka@northeastern.edu 📍 Boston, MA, USA 🔗 <https://katherine-atwell.github.io/>

🌐 github.com/katherine-atwell 🆔 <https://orcid.org/0000-0002-5066-5604>

EDUCATION

Ph.D., Computer Science Northeastern University GPA: 4.0	08/2020 – 05/2026 Boston, United States
B.S., Computer Science University of Maryland, Baltimore County	08/2016 – 05/2020 Baltimore, United States
B.S., Mathematics University of Maryland, Baltimore County	08/2016 – 05/2020 Baltimore, United States

PROFESSIONAL EXPERIENCE

Team Member Alexa Prize TaskBot Challenge Deployed task-oriented dialogue system to real Alexa users; won third place	01/2023 – 09/2023 Pittsburgh, United States
Software Engineering Intern Google, Inc. Developed a visualization tool for the Data Studio team	05/2019 – 08/2019 Mountain View, United States
Engineering Practicum Intern Google, Inc. Built an internal dashboarding tool for YouTube	05/2018 – 08/2018 Mountain View, United States
Research Intern (SPIRE-EIT REU) Iowa State University Designed the UX for new software for Traffic Incident Managers	06/2017 – 08/2017 Ames, United States

PUBLICATIONS

BASIL: Bayesian Assessment of Sycophancy in LLMs ArXiv preprint Katherine Atwell, Pedram Heydari, Anthony Sicilia, Malihe Alikhani	2026
Perspectives of Global Deaf and Hard-of-hearing Community Members on Sign Language Technologies ArXiv preprint Katherine Atwell, Saki Imai, Danielle Bragg, Malihe Alikhani	2026
MixDPO: Modeling Preference Strength for Pluralistic Alignment ArXiv preprint Saki Imai, Pedram Heydari, Anthony Sicilia, Asteria Kaeberlein, Katherine Atwell, Malihe Alikhani	2026
Measuring Bias and Agreement in Large Language Model Presupposition Judgments Findings of ACL 2025 Katherine Atwell, Mandy Simons, Malihe Alikhani	2025

Contextual ASR Error Handling with LLMs Augmentation for Goal-Oriented Conversational AI <i>COLING 2025 Industry Track</i> Yuya Asano, Sabit Hassan, Paras Sharma, Anthony Sicilia, Katherine Atwell , Diane Litman, Malihe Alikhani	2025
Generating Signed Language Instructions in Large-Scale Dialogue Systems <i>NAACL 2024 Industry Track</i> Mert Inan, Katherine Atwell , Anthony Sicilia, Lorna Quandt, Malihe Alikhani	2024
Studying and Mitigating Biases in Sign Language Understanding Models <i>EMNLP 2024</i> Katherine Atwell , Danielle Bragg, Malihe Alikhani	2024
The Importance of Including Signed Languages in Natural Language Processing <i>Sign Language Machine Translation, Springer Nature Switzerland</i> Kayo Yin, Katherine Atwell , Julie A Hochgesang, Malihe Alikhani	2024
Combining Discourse Coherence with Large Language Models for More Inclusive, Equitable, and Robust Task-Oriented Dialogue <i>COLING 2024</i> Katherine Atwell , Mert Inan, Anthony B Sicilia, Malihe Alikhani	2024
Dialogue with Robots: Proposals for Broadening Participation and Research in the SLIVAR Community <i>arXiv preprint arXiv:2404.01158</i> Casey Kennington, Malihe Alikhani, Heather Pon-Barry, Katherine Atwell , Yonatan Bisk, Daniel Fried, Felix Gervits, Zhao Han, Mert Inan, Michael Johnston, Raj Korpan, Diane Litman, Matthew Marge, Cynthia Matuszek, Ross Mead, Shiwali Mohan, Raymond Mooney, Natalie Parde, Jivko Sinapov, Angela Stewart, Matthew Stone, Stefanie Tellex, Tom Williams	2024
ISABEL: An inclusive and collaborative task-oriented dialogue system <i>Alexa Prize TaskBot Challenge 2 Proceedings</i> Anthony Sicilia, Yuya Asano, Katherine Atwell , Qi Cheng, Dipunj Gupta, Sabit Hassan, Mert Inan, Jennifer Nwogu, Paras Sharma, Malihe Alikhani.	2023
Multilingual Content Moderation: A Case Study on Reddit. <i>EACL 2023</i> Meng Ye, Karan Sikka, Katherine Atwell , Sabit Hassan, Ajay Divakaran, Malihe Alikhani	2023
How people talk about each other: Modeling Generalized Intergroup Bias and Emotion <i>EACL 2023</i> Venkata Govindarajan, Katherine Atwell , Barea Sinno, Malihe Alikhani, David Beaver, Junyi Jessy Li.	2023
APPDIA: A Discourse-aware Transformer-based Style Transfer Model for Offensive Social Media Conversations <i>COLING 2022</i> Katherine Atwell* , Sabit Hassan*, Malihe Alikhani	2022
The Role of Context and Uncertainty in Shallow Discourse Parsing <i>COLING 2022</i> Katherine Atwell , Remi Choi, Junyi Jessy Li, Malihe Alikhani	2022
PAC-Bayesian Domain Adaptation Bounds for Multiclass Learners <i>UAI 2022</i> Anthony Sicilia, Katherine Atwell , Malihe Alikhani, Seong Jae Hwang	2022
Political Ideology and Polarization: A Multi-dimensional Approach <i>NAACL 2022</i> Barea Sinno*, Bernardo Oviedo*, Katherine Atwell* , Malihe Alikhani, Junyi Jessy Li	2022

The Change that Matters in Discourse Parsing: Estimating the Impact of Domain Shift on Parser Error <i>Findings of ACL 2022</i> Katherine Atwell*, Anthony Sicilia*, Seong Jae Hwang, Malihe Alikhani	2022
Where Are We in Discourse Relation Recognition? <i>SIGDIAL 2021</i> Katherine Atwell, Junyi Jessy Li, Malihe Alikhani	2021
Developing an Optimized UI for Traffic Incident Managers <i>HFES 2018</i> Andrina Helgeson, Jamiahus Walton, Celia Loya, Christopher Kawell, Katherine Atwell , Quinn Monaghan, Lakshay Ahuja, Hesham Hassan, Stephen Gilbert, & Anuj Sharma	2018

TEACHING EXPERIENCE

Instructor of Record <i>Northeastern University</i> Teaching CS4120, an introductory computer science course for undergraduate students at Northeastern	01/2026 – Present Boston, US
Teaching Assistant <i>Northeastern University</i> Taught an NLP course and started an LLM track connecting principles learned in class to modern NLP systems	09/2025 – Present Boston, United States
Teaching Assistant <i>Northeastern University</i> Taught an unsupervised learning and data mining course	2024 Boston, United States
Teaching Assistant <i>University of Pittsburgh</i> Taught an AI ethics course	2021 Pittsburgh, United States
Teaching Assistant <i>University of Maryland, Baltimore County</i> Taught an object-oriented programming course	2019 – 2020 Baltimore, United States
Mathematics Tutor <i>University of Maryland, Baltimore County</i> Tutored for variety of math classes for the Learning Resource Center, including calculus, statistics, and real analysis	2017 – 2019 Baltimore, United States
Teaching Assistant <i>University of Maryland, Baltimore County</i> Taught Calculus I	2018 Baltimore, United States

INVITED TALKS

Prompt Design and Post-Training for LLMs <i>Northeastern University</i>	2024 Boston, United States
Responsible Artificial Intelligence and Marginalized Communities <i>Bilkent University</i>	2024 Ankara, Türkiye
Conversational AI Systems <i>Northeastern University</i>	2024 Boston, United States
Coreference <i>University of Pittsburgh</i>	2023 Pittsburgh, United States
APPDIA: A Discourse-aware Transformer-based Style Transfer Model for Offensive Social Media Conversations <i>The University of Texas at Austin</i>	2022 Austin, United States

AWARDS

Third Prize <i>Alexa Prize TaskBot Challenge</i>	2023
Best Paper <i>Association for Uncertainty in Artificial Intelligence</i>	2022
Travel Scholarship Recipient <i>CRA-W Scholarship for Early Career Workshop and the Graduate Cohort</i>	2022
Travel Scholarship Recipient <i>Grace Hopper Celebration</i>	2018
Meyerhoff Scholar <i>University of Maryland, Baltimore County</i>	2016

ACADEMIC SERVICE

Organizer <i>NLP for Positive Impact</i> Co-organized the fourth workshop for NLP for Positive Impact, which was co-located with ACL 2025	2025 Vienna, Austria
Organizer <i>PittInclusion</i> Organized the PittInclusion program to provide underrepresented students with research experience with faculty mentorship	2023 Pittsburgh, United States
Area Chair <i>ACL Rolling Review, CoNLL</i>	2024 – 2025
Ethics Reviewer <i>ACL Rolling Review</i>	2024 – present
Reviewer <i>ACL Rolling Review, EMNLP, ACL, NAACL, UAI, ACM FAccT, Journal of Educational Data Mining, INLG, AAAI</i>	
Program Committee <i>ACM FAccT, SIGDIAL, Workshop on Computational Approaches to Discourse</i>	
Graduate Application Reviewer <i>Northeastern University</i>	2024
Hiring Interview Scribe <i>Northeastern University</i>	2024