

MALIHE ALIKHANI

www.malihealikhani.com ◊ +1 (808)343-5754 ◊ m.alikhani@northeastern.edu

RESEARCH INTERESTS

I work towards developing safe and fair AI systems that enhance communication, decision-making, and knowledge-sharing across disciplines and populations. My work spans academia, public policy, and applied AI research, bridging the gap between theory, practice, and governance to ensure that AI serves society effectively and responsibly.

By integrating insights from cognitive science, neuroscience, philosophy, and social sciences with machine learning, I develop models that capture the richness of human interpretation and collaborative meaning construction.

EMPLOYMENT

Assistant Professor, Khoury College of Computer Science Fall 2023–Now
Northeastern University

Visiting Fellow at The Center on Regulation and Markets at Brookings February
2025–Now
American Association for the Advancement of Science

Science and Technology Congressional AI Fellow August 2024–February 2025
American Association for the Advancement of Science

Assistant Professor, Department of Computer Science Fall 2020–Fall 2023
University of Pittsburgh

Affiliations: Northeastern Ethics Institute, Northeastern Experiential AI Institute

EDUCATION

Ph.D. Computer Science 2015–2020
Rutgers University
Dissertation Title: Multimodal Communication: Coherence, Common Ground and Coordination
Dissertation committee: Matthew Stone (chair), Ani Nenkova, Gerard de Melo, Kostas Bekris

Graduate Certificate, Cognitive Science 2017–2019
Rutgers University
Thesis Project: Semantic Memory and Vague Categories in Communication

MSc Computer Science 2015–2017
Rutgers University

GRANTS AND SPONSORSHIP

PI, NSF, Designing Educational Tools for Deaf Students \$500,000	June 2024
Co-PI, DARPA, Modeling Friction in Human-AI Collaboration \$1,000,000	February 2024
Co-I, CDC, Modeling the Virality of Vaccine Misinformation \$1,000,000	January 2024
PI, Alexa Prize TaskBot Challenge Amazon Inc, \$500,000	January 2023
PI, NSF grant for hosting a workshop on multimodal interactive systems \$40,000	January 2023
PI, The Grate Moderators-Civil Sanctuary, DARPA, \$1,000,000	November 2021
Co-PI, Automatic Multimodal Affect Detection for Research and Clinical Use, NIH, \$1,500,000	September 2020
PI, Google Teaching Award for Diversity in AI Curriculum \$10,000	October 2021
PI, Google Explore CSR \$20,000	February 2020
PI, Pitt Momentum Teaming Grant Developing Multimodal Technology to Improve Deaf Students Access to STEM Education, \$50,000	February 2021
PI, PittCyber Teaming Grant Modeling the Effect of Disinformation on COVID-19 Vaccination Discourse, \$10,000	February 2021
PI, PittCyber Seed Grant Modeling Coordination in Political Discourse, \$5,000	February 2021
PI, Google ExploreCSR \$20,000	February 2021

SELECTED PUBLICATIONS

Books

- Statistics and Probability Common Core, Malihe Alikhani, Cliff Notes, 2015.

Book Chapters

- Sign Machine Translation, Malihe Alikhani, Julie A. Hochgesang, Springer. Editor: Andy Way
- Coherence and Collaboration, in Coherence in Discourse (to appear), Matthew Stone, Malihe Alikhani, Baber Khalid, Springer. Editor: Gerhard Preyer.
- Speech-Language Pathology as a Global Practice: Culture, Context, and Collaboration, to be published by Routledge by Malihe Alikhani, Sulare Telford, and Anthony Sicilia.

Selected Peer-reviewed Publications

1. Evaluating Theory of (an uncertain) Mind: Predicting the Uncertain Beliefs of Others in Conversation Forecasting, Anthony Sicilia and Malihe Alikhani, NAACL 2025.
2. Accounting for Sycophancy in Language Model Uncertainty Estimation, Anthony Sicilia, Mert Inan, Malihe Alikhani, NAACL 2025.
3. Coherence-Driven Multimodal Safety Dialogue with Active Learning for Embodied Agents, Sabit Hassan and Malihe Alikhani, AAMAS 2025.
4. Contextual ASR Error Handling with LLM Augmentation for Goal-Oriented Conversational AI, Yuya Asano, Anthiny Sicilia, Malihe Alikhani, COLING Industry Track 2025.
5. An Active Learning Framework for Inclusive Generation by Large Language Models, Sabit Hassan, Anthiny Sicilia, Malihe Alikhani, COLING 2025.
6. Active Learning for Robust and Representative LLM Generation in Safety-Critical Scenarios, Sabit Hassan, Anthony Sicilia, Malihe Alikhani, Workshop on Customizable NLP at EMNLP 2024. **(Outstanding Paper Award)**
7. Eliciting Uncertainty in Chain-of-Thought to Mitigate Bias against Forecasting Harmful User Behaviors. Anthiny Sicilia, Malihe Alikhani, 3rd Workshop on NLP for Positive Impact at EMNLP 2024. **(Outstanding Paper Award)**
8. Deal, or no deal (or who knows)? Forecasting Uncertainty in Conversations using Large Language Models, Anthony Sicilia, Hyunwoo Kim, Khyathi Raghavi Chandu, Malihe Alikhani, Jack Hessel, The 62nd Annual Meeting of the Association for Computational Linguistics (ACL) 2024.
9. Generating Signed Language Instructions in Large-Scale Dialogue Systems, Mert Inan, Katherine Atwell, Anthony Sicilia, Lorna Quandt, Malihe Alikhani, 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics.

10. HumBEL: A Human-in-the-Loop Approach for Evaluating Demographic Factors of Language Models in Human-Machine Conversations, Anthony Sicilia, Malihe ALikhani, The 18th Conference of the European Chapter of the Association for Computational Linguistics 2024.
11. A corpus of Persian literary text, Shahab Raji, Malihe Alikhani, Gerard de Melo, Matthew Stone, Journal of Language Resources and Evaluation, 2024.
12. Combining Discourse Coherence with Large Language Models for More Inclusive, Equitable, and Robust Task-Oriented Dialogue, Katherine Atwell, Mert Inan, Anthony B Sicilia, Malihe Alikhani, LREC-COLING 2024 - The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation.
13. Seeing Eye-to-Eye: Cross-Modal Coherence Relations Inform Eye-gaze Patterns During Comprehension and Production, Mert Inan, Malihe Alikhani, LREC-COLING 2024 - The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation.
14. Soda: Million-scale dialogue distillation with social commonsense contextualization, Hyunwoo Kim, Jack Hessel, Liwei Jiang, Ximing Lu, Youngjae Yu, Pei Zhou, Ronan Le Bras, Malihe Alikhani, Gunhee Kim, Maarten Sap, Yejin Choi, EMNLP 2023. **(Outstanding Paper Award)**
15. Learning Multimodal Cues of Children’s Uncertainty, Qi Cheng, Mert Inan, Rahma Mbarki, Grace Grmek, Theresa Choi, Yiming Sun, Kimele Persaud, Jenny Wang, Malihe Alikhani, Annual Meeting of the Special Interest Group on Discourse and Dialogue (Sigdial) 2023.
16. Image–text coherence and its implications for multimodal AI, Malihe Alikhani, Baber Khalid, Matthew Stone, Frontiers in Artificial Intelligence, 2023.
17. DisCGen: A Framework for Discourse-Informed Counterspeech Generation, Sabit Hassan, Malihe Alikhani, IJCNLP-AAACL 2023 : The 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics.
18. Learning To Generate Equitable Text in Dialogue from Biased Training Data, Anthony Sicilia, Malihe Alikhani, The 61st Annual Meeting of the Association for Computational Linguistics 2023.