Andrew Lee

Contact Information	**E-mail**: ajyl@umich.edu **Website**: https://ajyl.github.io

Research Interests

Natural Language Processing: Dialogue, Computational Social Science

Education

University of Michigan		Ann Arbor, MI
Ph.D. Candidate in Computer Science		2020 - Present
Advisor: Rada Mihalcea		
University of Michigan		Ann Arbor, MI
Master's in Computer Science		2015
GPA: 3.85/4.0 -- Summa Cum Laude		
Northwestern University		Evanston, IL
Bachelor of Science in Computer Science		2013
GPA: 3.61/4.0		

Publications

A. Lee, J. K. Kummerfeld, L. An, R. Mihalcea. 2021. Micromodels for Efficient, Explainable, and Reusable Systems: A Case Study on Mental Health. *Findings of Empirical Methods in Natural Language Processing (Findings of EMNLP)*.

S. Larson, A. Mahendran, J. J. Peper, C. Clarke, **A. Lee**, P. Hill, J. K. Kummerfeld, K. Leach, M. A. Laurenzano, L. Tang and J. Mars. 2019. An Evaluation for Intent Classification and Out-of-Scope Prediction. *Empirical Methods in Natural Language Processing (EMNLP)*.

S. Larson, A. Mahendran, **A. Lee**, J. K. Kummerfeld, P. Hill, M. Laurenzano, J. Hauswald, L. Tang, J. Mars. 2019. Outlier Detection for Improved Data Quality and Diversity in Dialog Systems. *North American Chapter of the Association for Computational Linguistics (NAACL)*.

Experience

Microsoft Research	Redmond, WA
Research Intern	May 2021 - August 2021
Advisor: Silviu-Petru Cucerzan	
Clinc, Inc.	Ann Arbor, MI
Core AI R&D - Senior Software Engineer, Team Lead	June 2019 - August 2020
Core AI R&D - Software Engineer	June 2017 - June 2019
Ford Motor Company	Dearborn, MI
Software Engineer	March 2016 - June 2017

Patents

Systems and methods for constructing an artificially diverse corpus of training data samples for training a contextually-biased model for a machine learning-based dialogue system.
A. Lee, S. Larson, C. Clarke, K. Leach, J. Kummerfeld, P. Hill, J. Hauswald, M. Laurenzano, L. Tang, J. Mars.
US Patent 10,796,104. 2020.

Systems and methods for automatically conguring training data for training machine learning models of a machine learning-based dialogue system including seeding training samples or curating a corpus of training data based on instances of training data identified as anomalous.
S. Larson, A. Mahendran, **A. Lee**, J. Kummerfeld, P. Hill, M. Laurenzano, J. Hauswald, L. Tang, J. Mars.
US Patent 10,679,150. 2020.

References   Available upon request.