Aida Ramezani

armzn@cs.toronto.edu

Education

2020 - present (PhD)	Computer Science at **University of Toronto**.
Supervisor:	Professor Yang Xu
Thesis focus:	Natural language processing for moral inference.
Research interests:	Morality in AI, AI safety, Natural Language Processing, Computational social science, Cultural analytics.
2016 - 2020 (B.Sc.)	Computer Engineering at **Sharif University of Technology**.
GPA:	18.92/20.0

Work Experience

Research intern at Microsoft + Nuance May 2023 - July 2023

- Improved NLP models for medical report summarization by integrating **generative language models**.
- Introduced a new dataset for boilerplate detection in medical reports.
- Developed and fine-tuned a pipeline for incorporating boilerplate detection into seq2seq **summarization models**.
- **Presented project findings** to technical audiences, showcasing significant improvements in summarization performance.
- Facilitated **communication with stakeholders** through multiple meetings, ensuring alignment on project milestones and goals.

Publications

[1] Aida Ramezani and Yang Xu. "Moral association graph: A cognitive model for moral inference". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 46. 2024.

[2] Aida Ramezani and Yang Xu. "Knowledge of cultural moral norms in large language models". In: *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics*. 2023.

[3] Aida Ramezani et al. "Evolution of moral semantics through metaphorization". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 44. 44. 2022.

[4] Aida Ramezani et al. "The emergence of moral foundations in child language development". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 44. 44. 2022.

[5] Aida Ramezani et al. "An unsupervised framework for tracing textual sources of moral change". In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021.

Manuscripts under review

[1] Aida Ramezani et al. *Evolution of the moral lexicon.*

[2] Aida Ramezani et al. *Quantifying the emergence of moral foundational lexicon in child language development.*

Awards

Cognitive Science Society Disciplinary Diversity & Integration Award	2024
Schwartz Reisman Institute for Technology and Society Graduate Affiliation.	2022 – 2023
Schwartz Reisman Institute for Technology and Society Graduate Fellowship.	2021 – 2022
University of Toronto, Recognition Of Excellence Award.	2020
Iran's National Elites Foundation: Recognized as elite member.	2016 – 2020

Academic research projects

Scalable inference of historical moral change August 2023 - Present

Developed a novel methodology based on **graph neural networks** and **language models** to identify historical cases of moralization in **large-scale** historical corpora and social media datasets. Incorporated **cognitive studies** of mental semantics and **psychological studies** of moralization in computational modelling.

Cultural moral knowledge in large language models Jul 2022 - Present

Proposed a novel framework for evaluating cultural moral knowledge within large language models. Identified and addressed significant moral biases and developed adaptive strategies to **align language models** with culturally diverse **human moral values**.

Moral language in child speech Jun 21 - Sep 2022

Introduced a novel methodology for **word-sense disambiguation** within the moral domain, enhancing computational text analysis of child language datasets, using expertise in **NLP** and insight from **moral psychology**. **Mentored** an undergraduate student and collaborated on interdisciplinary research. Currently in the process of publishing the findings to a high-profile interdisciplinary journal.

Evolution of moral semantics Mar 2021 - Sep 2022

Developed a computational framework for investigating the historical evolution of moral lexicon through **diachronic word embeddings**.

Tracing textual sources of moral change Sep 2020 - Jun 2021

Developed an **unsupervised probabilistic NLP** framework to quantify different influential sources on moral change in society. Worked with **large-scale** textual datasets of news articles and social media.

Skills

Machine learning	PyTorch, Weights & Biases, transformers, NumPy, Pandas, SciPy, Scikit-learn, PyG, RL4LM, JAX.
Data science	R, Data visualization.
Programming	Python, Linux, Java, LaTeX, Markdown, C, C++.
Mathematics & Statistics	Linear Algebra, Probability theory, Causal inference, Algebra, Calculus.
Languages	English (fluent), Persian (native), ASL (basic).

Talks and presentations

The emergence of the moral foundational lexicon in child language development
Moral Language Workshop, Institut Jean Nicod, Fall 2023

Machine inference of moralization across timescales
Morality Lab, Department of Psychology, University of Toronto, Fall 2023

Moral norm variation in large language models
ARIA 2023, University of Toronto, Fall 2023

Knowledge of cultural moral norms in large language models
Poster presentation, ACL 2023

The emergence of moral foundations in child language development
Oral presentation, Cogsci 2022

Evolution of moral semantics through metaphorization
Poster presentation, Cogsci 2022

Mentorship

University of Toronto, Winter 2024
Currently mentoring a computer science master's student working on moral inference in computer vision.

University of Toronto, Summer 2022
Mentored a computer science undergraduate student working on the dynamics of moral semantic change.

University of Toronto, Summer 2021
Mentored a computer science undergraduate student working on moral language in child speech.

Teaching

Neural Networks and Deep Learning, University of Toronto, Winter 2024

Natural language computing, University of Toronto, Winter 2023-2024

Computational linguistics, University of Toronto, Fall 2023

Computational models of semantic change, University of Toronto, Winter 2022

Introduction to artificial intelligence, University of Toronto, Winter 2021, Fall 2022

Foundations of Computer Science, University of Toronto, Fall 2021

Volunteering

CL Colloquium University of Toronto, Winter 2024
Co-organizing the Computational Linguistics speaker series.

PRISM University of Toronto, Winter 2024
Providing research mentorship to a group of 5 undergraduate computer science students.

Science Rendezvous University of Toronto, Spring 2024
Co-organizing a youth-outreach program for ethics of AI.

Science Rendezvous University of Toronto, Spring 2023
Co-organizing a youth-outreach program for the Turing test in computer vision.