Venkata S Govindarajan

DEPARTMENT OF LINGUISTICS,
THE UNIVERSITY OF TEXAS AT AUSTIN

he/him/his
venkatasg
gvenkatas
gvenkata1994@gmail.com

PERSONAL STATEMENT

I am a Ph.D. candidate at UT Austin studying intergroup bias in online communication. I have 5+ years of experience in **Natural Language Processing** (NLP) including large language models, and am interested in **Machine Learning** or **Data Science** roles *starting in May 2024*.

EDUCATION

The University of Texas at Austin — 2019–2024
Ph.D. Computational Linguistics

University of Rochester — 2017–2019
M.S. Computational Linguistics

Indian Institute of Technology Madras — 2012–2017
B.Tech & M.Tech Biological Engineering

EMPLOYMENT

The New York Times — Summer 2023
Data Science Intern
Built a framework using mixed effects linear modeling for offline evaluation of novel user and article features to inform model building in algorithmic recommendations. Delivered potential features that could **boost engagement up-to 2%** while diversifying recommendations.

Amazon — Summer 2021
Alexa Applied Scientist Intern
Implemented an unsupervised method for detecting data drift in NLU models, and validated the approach on simulated and customer data. Received return internship offer for summer 2022.

SELECT PUBLICATIONS

1. "Lil-Bevo: Explorations of Strategies for Training Language Models in More Humanlike Ways" (Dec. 2023). In: *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*. Singapore: Association for Computational Linguistics (ACL).

2. "Counterfactual Probing for the Influence of Affect and Specificity on Intergroup Bias" (July 2023). In: *Findings of the ACL: ACL 2023*. Toronto, Canada.

3. "How people talk about each other: Modeling Generalized Intergroup Bias and Emotion" (May 2023). In: *Proceedings of the 17th Conference of the European Chapter of the ACL (EACL 2023)*. Dubrovnik, Croatia.

4. "Help! Need Advice on Identifying Advice" (Nov. 2020). In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

5. "Decomposing Generalization: Models of Generic , Habitual, and Episodic Statements" (2019). In: *Transactions of the ACL (TACL)* 7.

SKILLS

Programming Languages: Python, Swift, Javascript, R, SQL, Julia, LISP

Tools & Frameworks: pyTorch, Transformers, Tensorflow, Keras, scikit-learn, statsmodels, SciPy, Pandas, SwiftUI, CoreML, BigQuery, lme4

Languages: English (native), Tamil (native), Hindi (intermediate)

RELEVANT COURSEWORK

Machine Learning, Natural Language Processing, Natural Language Generation & Discourse Processing, Logical Foundations of Artificial Intelligence, Statistical Speech Processing

APPS

DeTeXt: I built an open source app for iOS, iPadOS and macOS that predicts the best LaTeX commands corresponding to hand-drawn symbols using deep convolutional neural networks. Built using SwiftUI, Combine, PencilKit and CoreML, the app has **over 10,000 downloads**.

PROFESSIONAL SERVICE

Organizing Committee

South by Semantics Workshop 2022 –23.
Texas Linguistics Society (TLS) Conference 2021–22.

Reviewer

ACL ARR 2023, EMNLP 2023, ACL 2023, *SEM 2023–24, SIGDIAL 2022–23

AWARDS

NASSLI Student Grant (800 USD)	Summer 2022
COLA Supplemental Graduate School Fellowship (5000 USD)	Spring 2020
Silver medal at International Genetically Engineered Machine (iGEM)	Fall 2016
Indian Biological Engineering Competition (iBEC) grant (15,000 USD)	Fall 2016
National BIRAC-IdeaThon on Antimicrobial Resistance Finalist	Fall 2016
Second runner up in 3M-CII Young Innovators Challenge	Spring 2015