Venkata S Govindarajan

DRIVEN BY LEARNING & BUILDING

GitHub LinkedIn Google Scholar
venkat@venkatasg.net +1(512)568-8775
Ithaca, NY 14850, USA (Open to relocation)

EDUCATION

The University of Texas at Austin (*Ph.D. Computational Linguistics*) 2019–2024
GPA: 3.92/4

University of Rochester (*M.S. Computational Linguistics*) 2017–2019
GPA: 3.75/4

Indian Institute of Technology Madras (*B.Tech & M.Tech Bioengineering*) 2012–2017
GPA: 8.68/10

EMPLOYMENT

Ithaca College August 2024–Present
Assistant Professor of Computer Science Ithaca, NY
Developed a modern upper-level course on Natural Language Processing (NLP), covering both classical NLP and **Large Language Models (LLM)**, including architectures, training pipelines and cutting-edge applications such as retrieval-augmented generation, reasoning, tool-calling and agents.

The New York Times June–August 2023
Data Science Intern Austin, TX
Developed a mixed-effects linear modeling framework (on Google Cloud & Vertex AI) to assess how user features, article attributes, and LLM embeddings influence reader engagement with suggested articles, discovering potential features that could **boost engagement up-to 2%** while diversifying recommendations.

Amazon May–August 2021
Alexa Applied Scientist Intern Austin, TX
Implemented an unsupervised latent factor graphical model to detect data drift — where live data diverges from training data — in deployed NLU systems; validated the approach on simulated customer datasets.

The University of Texas at Austin August 2019–May 2024
Graduate Research Assistant Austin, TX
Lead NLP research on intergroup bias and advice-giving in online discourse. Built pipelines to explore inductive biases in **low-resource LM training** and counterfactual probes to analyze multilingual representations.

APPS & PROJECTS

diversity: Contributed to an open-source package for evaluating the diversity of LLM outputs; I implemented batching and performance optimizations that made evaluations up to **10x faster**.

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

SELECT PUBLICATIONS

1. "Measuring Lexical Diversity of Synthetic Data Generated through Fine-Grained Persona Prompting" (Nov. 2025). In: *To appear in Findings of the Association for Computational Linguistics: EMNLP 2025*.

2. "Using Language Models for Music Recommendations with Natural-Language Profiles" (Sept. 2025). In: *Proceedings of the 3rd Music Recommender Systems Workshop (MuRS 2025)*. Prague: 19th ACM Conference on Recommender Systems (RecSys 2025).

3. "Do *they* mean 'us'? Interpreting Referring Expression variation under Intergroup Bias" (Nov. 2024). In: *Findings of the Association for Computational Linguistics: EMNLP 2024*. Miami.

4. "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).

5. "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.

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

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

SKILLS

Programming Languages: Python, Swift, Racket, Javascript, Typescript, R, SQL, LISP, C, C++

Tools & Frameworks: PyTorch, Huggingface, Tensorflow, Keras, scikit-learn, statsmodels, SciPy, Pandas, SwiftUI, CoreML, Google Cloud Platform (GCP), BigQuery, lme4, CUDA, Metal/MLX, Weights & Biases

PROFESSIONAL SERVICE

Organizing Committee

South by Semantics Workshop 2022–24.
Texas Linguistics Society (TLS) Conference 2021–22.

Reviewer

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