Venkata S Govindarajan

DEPARTMENT OF COMPUTER SCIENCE,
ITHACA COLLEGE

🌐 venkatasg.net
✉ venkat@venkatasg.net
🎓 Google Scholar
in LinkedIn

RESEARCH INTERESTS

Natural Language Processing, Computational Linguistics & Social Science

---

EMPLOYMENT

Ithaca College   August 2024–Present
Assistant Professor of Computer Science

The New York Times   June–August 2023
Data Science Intern

Amazon   May–August 2021
Alexa Applied Scientist Intern

EDUCATION

The University of Texas at Austin   2019–2024
Ph.D. Computational Linguistics   GPA: 3.92/4
Thesis: Modeling Intergroup Bias in Online Conversation.
Committee: Jessy Li, David Beaver, Kyle Mahowald & Malihe Alikhani

University of Rochester   2017–2019
M.S. Computational Linguistics   GPA: 3.75/4
Thesis: Models of Generic, Habitual and Episodic Statements.
Advisor: Aaron Steven White

Indian Institute of Technology Madras   2012–2017
B.Tech & M.Tech Biological Engineering   GPA: 8.68/10

PUBLICATIONS

1. Gauri Kambhatla, Chantal Shaib, and **Venkata Govindarajan** (May 2025). "Measuring diversity of synthetic prompts and data generated with fine-grained persona prompting". arXiv: 2505.17390.

2. **Venkata S Govindarajan**, Matianyu Zang, Kyle Mahowald, David Beaver, and Junyi Jessy Li (Nov. 2024). "Do *they* mean 'us'? Interpreting Referring Expression variation under Intergroup Bias". In: *Findings of the Association for Computational Linguistics: EMNLP 2024*. Ed. by Yaser Al-

Onaizan, Mohit Bansal, and Yun-Nung Chen. Miami, Florida, USA: Association for Computational Linguistics, pp. 9772–9785.

3. Anirudh Srinivasan, **Venkata S Govindarajan**, and Kyle Mahowald (Dec. 2023). "Counterfactually Probing Language Identity in Multilingual Models". In: *Proceedings of the 3rd Workshop on Multi-lingual Representation Learning (MRL)*. Ed. by Duygu Ataman. Singapore: Association for Computational Linguistics, pp. 24–36.

4. **Venkata S Govindarajan**, Juan Diego Rodriguez, Kaj Bostrom, and Kyle Mahowald (Dec. 2023). "Lil-Bevo: Explorations of Strategies for Training Language Models in More Humanlike Ways". In: *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*. Ed. by Alex Warstadt, Aaron Mueller, Leshem Choshen, Ethan Wilcox, Chengxu Zhuang, Juan Ciro, Rafael Mosquera, Bhargavi Paranjabe, Adina Williams, Tal Linzen, and Ryan Cotterell. Singapore: Association for Computational Linguistics, pp. 280–288.

5. **Venkata S Govindarajan**, David Beaver, Kyle Mahowald, and Junyi Jessy Li (July 2023). "Counterfactual Probing for the Influence of Affect and Specificity on Intergroup Bias". In: *Findings of the Association for Computational Linguistics: ACL 2023*. Toronto, Canada: ACL, pp. 12853–12862.

6. **Venkata S Govindarajan**, Katherine Atwell, Barea Sinno, Malihe Alikhani, David Beaver, and Junyi Jessy Li (May 2023). "How people talk about each other: Modeling Generalized Intergroup Bias and Emotion". In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. Dubrovnik, Croatia: ACL, pp. 2488–2498.

7. Venelin Kovatchev, Trina Chatterjee, **Venkata S Govindarajan**, Jifan Chen, Eunsol Choi, Gabriella Chronis, Anubrata Das, Katrin Erk, Matthew Lease, Junyi Jessy Li, Yating Wu, and Kyle Mahowald (July 2022). "longhorns at DADC 2022: How many linguists does it take to fool a Question Answering model? A systematic approach to adversarial attacks." In: *Proceedings of the First Workshop on Dynamic Adversarial Data Collection*. Seattle, WA: ACL, pp. 41–52.

8. **Venkata S Govindarajan**, Benjamin Chen, Rebecca Warholic, Katrin Erk, and Junyi Jessy Li (Nov. 2020). "Help! Need Advice on Identifying Advice". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: ACL, pp. 5295–5306.

9. Aaron Steven White, Elias Stengel-Eskin, Siddharth Vashishtha, **Venkata S Govindarajan**, Dee Ann Reisinger, Tim Vieira, Keisuke Sakaguchi, Sheng Zhang, Francis Ferraro, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme (May 2020). "The Universal Decompositional Semantics Dataset and Decomp Toolkit". In: *Proceedings of The 12th Language Resources and Evaluation Conference (LREC)*. Marseille, France: European Language Resources Association, pp. 5698–5707.

10. **Venkata S Govindarajan**, Benjamin Van Durme, and Aaron Steven White (2019). "Decomposing Generalization: Models of Generic , Habitual, and Episodic Statements". In: *Transactions of the Association for Computational Linguistics (TACL)* 7, pp. 501–517.

TALKS

1. "Diving into LLMs like ChatGPT: How do they work?" (May 2025). Hopshire's Science & Suds (Public outreach). Ithaca, NY.

2. "Computational Linguistic Models of Social Identity" (Feb. 2024). Invited talk. Ithaca College.

3. "Computational Models of Social Meaning" (Jan. 2024). Invited talk. Georgetown University.

4. "Modeling Generalized Intergroup Bias" (May 2023). Conference talk at EACL 2023.

5. "Help! Need Advice on Identifying Advice" (Nov. 2020). Conference talk at EMNLP 2020.

6. "Decomposing Generalization" (July 2020). Conference talk at ACL 2020.

TEACHING & MENTORSHIP

Teaching

Natural Language Processing   Spring 2025
Principal Instructor at Ithaca College   COMP490

Programming languages   Fall 2024
Principal Instructor at Ithaca College   COMP321

Principles of Computer Science   Fall 2024, Spring2025
Principal Instructor at Ithaca College   COMP171

Language and Computers   Summer 2022
Principal Instructor at UT Austin   LINS313

Teaching Assistant

- Machine Learning Toolbox for Text Analysis (LIN373N)   Spring 2021
- Analyzing Linguistic Data and Programming for Linguists (LIN350) Spring 2020
- Introduction to Computational Linguistics (LIN353C)   Fall 2019
- Introduction to Computational Linguistics   Fall 2018
- Data Structures and Algorithms for Biology   Fall 2016

Mentorship

Dobbs Conrad (undergraduate at Ithaca College)   Summer 2025–present
Constrained LLM generation for aiding language learning

Suryash Malviya (undergraduate at Ithaca College)   Spring 2025–present
GPU Programming, NPC dialogue synthesis with LLMs

SKILLS

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

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

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

PROFESSIONAL SERVICE

NASSLLI summer course   Summer 2025

I designed and taught a week-long summer course on Natural Language Processing & Computational Social Science (CSS) with my colleague Laura Biester at NASSLLI 2025, held at University of Washington, Seattle. The course focused on bridging theoretical concepts and engineering principles in NLP towards writing research papers in CSS, using insights from our past papers.

Tech Innovation Summer Academy (TISA)   Summer 2025

TISA is a one-week summer program that introduces students to computer science with a focus on understanding through building applications and tools. I served on the organizing committee helping with planning and organization, and designed course materials for sessions on Machine Learning and Artificial Intelligence.

South by Semantics Workshop   2022–24

I co-organized a series of talks on semantics, computation and philosophy of language. Our goal was to provide a forum for emerging scholars with UT Austin's rich community of language researchers, and to foster collaboration and dialogue through social events organized with the talks.

Texas Linguistics Society (TLS) Conference   2021–22

TLS is an annual student-run academic conference by the Department of Linguistics at UT Austin which includes presentations in all areas of linguistics. I served on the organizing committee and was involved in reviewing and assigning submissions for review, website design, scheduling, logistics for a fully virtual conference, and chairing keynote sessions and talks.

Reviewer
  ACL ARR 2023–present, EMNLP 2023–24, ACL 2023, *SEM 2023–25, SIGDIAL 2022–23

Volunteer
  Underpaid@UT (2023-24)

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 iOS, iPadOS and macOS that predicts the best LaTeX commands corresponding to hand-drawn symbols using deep neural networks. Built using SwiftUI, Combine, PencilKit and CoreML, the app has over 10,000 downloads.

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