

Venkata S Govindarajan

DEPARTMENT OF COMPUTER SCIENCE,
ITHACA COLLEGE

 venkatasg.net
 venkat@venkatasg.net
 [Google Scholar](#)
 [LinkedIn](#)

RESEARCH INTERESTS

Computational Linguistics, Computational Social Science, Natural Language Processing

EMPLOYMENT

Ithaca College <i>Assistant Professor in Computer Science</i>	August 2024–Present
The New York Times <i>Data Science Intern</i>	June–August 2023
Amazon <i>Alexa Applied Scientist Intern</i>	May–August 2021

EDUCATION

The University of Texas at Austin <i>Ph.D. Computational Linguistics</i> <i>Thesis: Modeling Intergroup Bias in Online Conversation.</i> <i>Committee: Jessy Li, David Beaver, Kyle Mahowald & Malihe Alikhani</i>	2019–2024 GPA: 3.92 / 4
University of Rochester <i>M.S. Computational Linguistics</i> <i>Thesis: Models of Generic, Habitual and Episodic Statements.</i> <i>Advisor: Aaron Steven White</i>	2017–2019 GPA: 3.75 / 4
Indian Institute of Technology Madras <i>B.Tech & M.Tech Biological Engineering</i>	2012–2017 GPA: 8.68 / 10

TEACHING

<i>Principal Instructor</i>	
Software Engineering <i>Ithaca College</i>	Spring 2026 COMP345

Natural Language Processing <i>Ithaca College</i>	Spring 2025 COMP490
Programming languages <i>Ithaca College</i>	Fall 2024 COMP321
Principles of Computer Science <i>Ithaca College</i>	Fall 2024/25, Spring 2025/26 COMP171
Language and Computers <i>UT Austin</i>	Summer 2022 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

UNDERGRADUATE MENTORSHIP

Dobbs Conrad Spring 2026
Constrained LLM generation for aiding language learning

Paul Gagliano, Griffin Homan, Cassandra Raineault May–Sept 2025
Music Recommendation with LLMs & Natural Language Profiles

Advised senior undergrads with my colleague Doug Turnbull exploring the impact of using LLMs under various settings for music recommendation, with a focus on their recommendation quality for long-tail artists. Culminated in a [student-lead paper and talk at 3rd Music Recommender Systems Workshop](#) (MuRS) co-located with 19th ACM Conference on Recommender Systems (RecSys 2025)

Suryash Malviya Spring 2025
GPU Programming

I advised one of my students on an independent research study learning GPU programming in numba-cuda and Apple’s Metal/Accelerate frameworks, culminating in an [approved PR](#) to the mlx repo that lead to 19% speed-up in complex matrix multiplications.

Intergroup Bias dynamics & discourse structure

Collaborated with, and mentored an undergraduate student from Brown who worked under my advisor as part of a summer research program. We built annotation interfaces, crowd-sourced annotations on a challenging dataset of interpersonal online language use, validated annotations, and fine-tuned LMs on our annotations to bootstrap large-scale labelling for insights into social language use. Lead to a [publication and presentation at EMNLP 2024](#).

PUBLICATIONS

1. **Venkata S Govindarajan** and Laura Biester (July 2026). “[Dark & Stormy: Modeling Humor in the Worst Sentences Ever Written](#)”. In: *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*. San Diego, USA: Association for Computational Linguistics.
2. Chantal Shaib, **Venkata S Govindarajan**, Joe Barrow, Jiuding Sun, Alexa Siu, Byron C Wallace, and Ani Nenkova (Dec. 2025). “[Standardizing the Measurement of Text Diversity: A Tool and Comparative Analysis](#)”. In: *Proceedings of The 14th International Joint Conference on Natural Language Processing and The 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics: System Demonstrations*. Ed. by Xuebo Liu and Ayu Purwarianti. Mumbai, India: Association for Computational Linguistics, pp. 36–46.
3. Gauri Kambhatla, Chantal Shaib, and **Venkata S Govindarajan** (Nov. 2025). “[Measuring Lexical Diversity of Synthetic Data Generated through Fine-Grained Persona Prompting](#)”. In: *Findings of the Association for Computational Linguistics: EMNLP 2025*. Suzhou, China.
4. Paul Gagliano, Griffin Homan, Douglas Turnbull, and **Venkata S Govindarajan** (Sept. 2025). “[Using Language Models for Music Recommendations with Natural-Language Profiles](#)”. In: *Proceedings of the 3rd Music Recommender Systems Workshop (MuRS 2025)*. Prague: 19th ACM Conference on Recommender Systems (RecSys 2025).
5. **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.
6. 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.
7. **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 Lan-*

guage 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.

8. **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.
9. **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.
10. 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.
11. **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.
12. 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.
13. **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. “Dark & Stormy: Bad Humor in the Worst Sentences Ever Written” (Apr. 2026). [Ithaca Public Works](#). Ithaca, NY.
2. “Diving into LLMs like ChatGPT: How do they work?” (May 2025). [Hopshire’s Science & Suds](#). Ithaca, NY.
3. “Computational Linguistic Models of Social Identity” (Feb. 2024). Invited talk. Ithaca College.
4. “Computational Models of Social Meaning” (Jan. 2024). Invited talk. Georgetown University.
5. “[Modeling Generalized Intergroup Bias](#)” (May 2023). Conference talk at EACL 2023.

6. “[Help! Need Advice on Identifying Advice](#)” (Nov. 2020). Conference talk at EMNLP 2020.
7. “Decomposing Generalization” (July 2020). Conference talk at ACL 2020.

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 organized by Ithaca College 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

APPS & PROJECTS

[fizzbuzz-bench](#): Designed and released an LLM benchmark that evaluates turn-based gameplay performance on FizzBuzz with configurable rules, measuring arithmetic and counting accuracy, long-horizon multi-turn reasoning, and generalization beyond training data.

TypeProof: Built with Google's Gemini CLI, TypeProof is a web app that helps you proof any typeface from the Google Fonts library using [Jonathan Hoefler's design-motivated typeproofs](#) that eschew pangrams.

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