

# Tushaar Gangavarapu

456 Gates Hall  
Cornell University  
Ithaca, NY

TG352@cornell.edu  
<https://tushaargvs.github.io>

## Education

- 2022–2024 MS in Computer Science, Cornell University  
2023–24: *Linear recurrent models for language modeling*  
Committee: Alexander “Sasha” Rush (chair) and David Bindel  
2022–24: *Forecasting derailment in online conversations*.  
Advisors: Cristian Danescu-Niculescu-Mizil (2022–24) and Lillian Lee (2022–23)
- 2015–2019 BTech in Information Technology, NIT Karnataka  
Thesis: *Psychological and behavioral traits in social-media language*  
Committee: Ram Mohana Reddy Guddeti (chair), Sowmya Kamath S, Nagamma Patil, Biju R Mohan

## Experience

- Spring 2025 Co-instructor (with Karthik Sridharan), Cornell University  
CS 3780/5780: Intro to Machine Learning
- Summer 2024 Research Intern, Cornell Tech. Advisor: Sasha Rush  
Worked on understanding the role of recurrence in hybrid models using sparse autoencoders
- Summer 2023 Research Intern, Cornell Tech. Advisor: Sasha Rush  
Worked on authorship identification in cross-domain settings
- 06/2019–08/2022 Applied Scientist (2021–2022), Research Engineer (2019–2021), Amazon  
Worked on automated algorithms to enhance user engagement in ebooks (Research published at AMLC 2022)
- Summer 2018 Software Intern, Kindle Create, Amazon

## Awards and honors

- 2022–23, 2023–24 Cornell Bowers CIS Best TA Award  
CS 4740 (NLP), CS 4300 (Language and Information)
- 2018–2019 Huawei National Scholarship for Academic Excellence
- 2013–2015 National Higher Secondary Education Scholarship  
(National rank: 10)
- 2012 South-Indian Mathematics Olympiad (national rank: 32)

## Publications

Google Scholar: [https://scholar.google.com/citations?user=C7v\\_cA8AAAAJ](https://scholar.google.com/citations?user=C7v_cA8AAAAJ).

**Current research interests.** NLP → LLMs → alternate-attention architectures → compute-efficient models (e.g., linear RNNs), interpretability (for model design)

- [1] Junxiong Wang, Tushaar Gangavarapu, Jing Nathan Yan, and Alexander M. Rush. MambaByte: Token-free Selective State Space Model, August 2024. URL <http://arxiv.org/abs/2401.13660>. arXiv:2401.13660
- [2] Johannes Knittel, Tushaar Gangavarapu, Hendrik Strobelt, and Hanspeter Pfister. GPT-2 Through the Lens of Vector Symbolic Architectures. October 2024. URL <https://neurips.cc/virtual/2024/workshop/84704>
- [3] Yann Hicke, Abhishek Masand, Wentao Guo, and Tushaar Gangavarapu. Assessing the efficacy of large language models in generating accurate teacher responses. In Ekaterina Kochmar, Jill Burstein, Andrea Horbach, Ronja Laarmann-Quante, Nitin Madnani, Anaïs Tack, Victoria Yaneva, Zheng Yuan, and Torsten Zesch, editors, *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 745–755, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.bea-1.60. URL <https://aclanthology.org/2023.bea-1.60>
- [4] Tushaar Gangavarapu and Sriraghavendra Ramaswamy. Alexa, stop reading the references: Enhancing the reading experience in Kindle eBooks. In *Proceedings of the Amazon Machine Learning Conference*, Seattle, WA, 2022a. Amazon
- [5] Tushaar Gangavarapu and Sriraghavendra Ramaswamy. A figure is worth a thousand words, but where are the words?: Enhancing image experience in Kindle eBooks. In *Proceedings of the Amazon Machine Learning Conference*, Seattle, WA, 2022b. Amazon
- [6] Tushaar Gangavarapu, Gokul S Krishnan, Sowmya Kamath S, and Jayakumar Jeganathan. FarSight: Long-Term Disease Prediction Using Unstructured Clinical Nursing Notes. *IEEE Transactions on Emerging Topics in Computing*, 9(3):1151–1169, July 2021. ISSN 2168-6750. doi: 10.1109/TETC.2020.2975251. URL <https://ieeexplore.ieee.org/document/9007352>. Conference Name: IEEE Transactions on Emerging Topics in Computing
- [7] Veena Mayya, Sowmya Kamath S., Gokul S. Krishnan, and Tushaar Gangavarapu. Multi-channel, convolutional attention based neural model for automated diagnostic coding of unstructured patient discharge summaries. *Future Generation Computer Systems*, 118: 374–391, May 2021. ISSN 0167-739X. doi: 10.1016/j.future.2021.01.013. URL <https://www.sciencedirect.com/science/article/pii/S0167739X21000236>
- [8] Tushaar Gangavarapu, Aditya Jayasimha, Gokul S. Krishnan, and Sowmya Kamath S. Predicting ICD-9 code groups with fuzzy similarity based supervised multi-label classification of unstructured clinical nursing notes. *Knowledge-Based Systems*, 190:105321, February 2020b. ISSN 0950-7051. doi: 10.1016/j.knosys.2019.105321. URL <https://www.sciencedirect.com/science/article/pii/S0950705119305982>
- [9] Aditya Jayasimha, Tushaar Gangavarapu, S. Sowmya Kamath, and Gokul S. Krishnan. Deep Neural Learning for Automated Diagnostic Code Group Prediction Using Unstructured Nursing Notes. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, CoDS COMAD 2020, pages 152–160, New York, NY, USA, January 2020. Association for Computing Machinery. ISBN 978-1-4503-7738-6. doi: 10.1145/3371158.3371176. URL <https://dl.acm.org/doi/10.1145/3371158.3371176>

- [10] Tushaar Gangavarapu, C. D. Jaidhar, and Bhabesh Chanduka. Applicability of machine learning in spam and phishing email filtering: review and approaches. *Artificial Intelligence Review*, 53(7):5019–5081, October 2020a. ISSN 1573-7462. doi: 10.1007/s10462-020-09814-9. URL <https://doi.org/10.1007/s10462-020-09814-9>
- [11] Tushaar Gangavarapu, Gokul S Krishnan, and Sowmya Kamath S. Coherence-based Modeling of Clinical Concepts Inferred from Heterogeneous Clinical Notes for ICU Patient Risk Stratification. In Mohit Bansal and Aline Villavicencio, editors, *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 1012–1022, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/K19-1095. URL <https://aclanthology.org/K19-1095>
- [12] Tushaar Gangavarapu and Nagamma Patil. A novel filter–wrapper hybrid greedy ensemble approach optimized using the genetic algorithm to reduce the dimensionality of high-dimensional biomedical datasets. *Applied Soft Computing*, 81:105538, August 2019. ISSN 1568-4946. doi: 10.1016/j.asoc.2019.105538. URL <https://www.sciencedirect.com/science/article/pii/S156849461930314X>

## Invited talks, lectures, etc.

- 2024                      Seagull LM: Generating humorous captions from scene descriptions  
Lecture for CS 4740/5740: Natural Language Processing, Cornell University
- 2024                      Tension with a chance of personal attack!  
Guest lecture for UNILWYL 1405, Cornell University
- 2024                      Gradient-based optimization and automatic differentiation  
Lecture for CS 4740/5740: Natural Language Processing, Cornell University  
Lecture notes: <https://github.com/TushaarGVS/backprop-lecture-notes-CS-4740>
- 2020                      Learning to Predict: Tree-based Classification  
Guest lecture at the Machine Learning University (MLU), Amazon  
Notes: <https://tushaargvs.github.io/assets/teaching/dt-notes-2020.pdf>
- 2020                      Cognitive and Affective Assessments in Game-based Simulations  
Invited talk at the Dept. of Information Technology, NITK
- 2020                      Greedy Evolutionary Feature Selection for Biomedical Data  
Invited talk at the Dept. of Information Technology, NITK
- 2019                      On the Convergence of HPC and Machine Intelligence  
Invited talk at High Performance Computing and Applications (HPCA)
- 2019                      Exploring Latent Human Traits Through Social Media Modeling  
Guest lecture at the Dept. of Information Technology, NITK
- 2019                      Game-based Learning and Assessment: A Case Study of a Mobile-VR Game.  
Guest lecture at the Dept. of Information Technology, NITK
- 2019                      Building Predictive Applications Using Social Media Digital Footprints.  
Invited talk at the Workshop on Predictive Analytics and Applications (PAA)

## Teaching

Spring 2025	Co-instructor, CS 3780/5780: Intro to Machine Learning Cornell University
Fall 2024	Head TA, CS 4701: Practicum in AI (mentored 6 project teams) Cornell University
Spring 2023, Spring 2024	Head TA, CS/INFO 4300: Language and Information Cornell University
Fall 2022, Fall 2023	Head TA (Fall 2023), Grad TA (Fall 2022), CS 4740/5740: NLP Cornell University

## Service

2024	Reviewer, EMNLP
2022	Reviewer, Amazon Machine Learning Conference (AMLC): Healthcare Informatics and NLP track
2021–2022	Organizer, Kindle algorithms weekly research meetings, Amazon

(Last compiled: 11/10/2024. Template inspired from [Chris Manning](#).)