

# Vivek Gupta

Machine Learning & Natural Language Processing

CONTACT INFORMATION	The Preserve Apartment 531 South 900 East, Apt. A8, Salt Lake City, UT 84102 United States of America	<a href="https://github.com/vgupta123">vgupta123.github.io</a> <a href="mailto:keviv9@gmail.com">keviv9@gmail.com</a> +1-801-558-7272 <a href="#">Google Scholar</a>
RESEARCH INTERESTS	Natural Language Inference, Representation Learning, Extreme Multi-Label Learning, Ethics in Artificial Intelligence	
CURRENT POSITION	<b>School of Computing, University of Utah</b> <i>Graduate Research Assistant</i>	Aug 2018 - Present Advisor: Prof. Vivek Srikumar
EDUCATION	<b>School of Computing, University of Utah</b> PhD, Computer Science, School of Computing Advisor: Prof. Vivek Srikumar (CGPA: 3.9/4)	Aug 2018 - Present
	<b>Indian Institute of Technology, Kanpur</b> M. Tech, Computer Science and Engineering Advisor: Prof. Harish Karnick (CGPA: 9.3/10)	July 2015 - May 2016
	<b>Indian Institute of Technology, Kanpur</b> B. Tech, Computer Science and Engineering (CGPA: 7.5/10)	July 2011 - July 2015
PREVIOUS EMPLOYMENT	<b>Microsoft Research Lab, India</b> <i>Research Fellow, Machine Learning and Natural Language Application</i> Dr. Praneeth Netrapalli & Dr. Prateek Jain	Oct 2016 - Aug 2018 Dr. Nagarajan Natarajan,
SCHOLASTIC ACHIEVEMENTS	One of the recipients of the <a href="#">Bloomberg Data Science Fellowship</a> 2021-2022. Selected to attend Virtual-HLF 2020, MLSS 2018 Madrid, MLSS 2020 London. Selected among the top 7 finalist in <a href="#">Ericson's Innovation Awards, India</a> in 2016 Selected in top 100 all over India in <b>Telnor, Internet for all challenge</b> by Unisys in 2015 Secured <b>All India Rank 183</b> in <b>IIT-JEE 2011</b> amongst more than 500,000 candidates Qualified for <b>Kishore Vaigyanik Protsahan Yojana (KVPY) scholarship, 2011</b> Selected among top 1% of students, <b>Indian National Physics Olympiads, 2011</b> Selected among top 1% of students, <b>Indian National Chemistry Olympiads, 2011</b>	
PUBLICATIONS	<b>Gupta, V.</b> , Zhang, S., Vempala, A., He, Y., Choji, T., Srikumar V., <i>Right for the Right Reason: Evidence Extraction for Trustworthy Tabular Reasoning</i> , <a href="#">ACL 2022</a> [Preprint] <b>Gupta, V.</b> , Bhat, R., Ghosal, A., Srivastava, M., Singh, M., Srikumar V., <i>Is My Model Using The Right Evidence? Systematic Probes for Examining Evidence-Based Tabular Reasoning</i> , <a href="#">TACL 2022</a> [Preprint] <b>Gupta, V.</b> , Shrivastava, A., Sagar, A., Aghajanyan, A., Savenkov. D., <i>RetroNLU: Retrieval Augmented Task Oriented Semantic Parsing</i> , under review [Preprint] Jain, N., <b>Gupta, V.</b> , Rai, A., Kumar, G., <i>TabPert: An Effective Platform for Tabular Perturbation</i> <a href="#">EMNLP 2021</a> , Demo track [Paper] [Project] <b>Gupta, V.</b> , Neeraja, J., Srikumar, V. <i>Incorporating External Knowledge to Enhance Tabular Reasoning</i> , <a href="#">NAACL 2021</a> [Paper] [Project] <b>Gupta, V.</b> , Mehta, M., Nokhiz, P., Srikumar, V. <i>InfoTabS: Inference on Tables as Semi-structured Data</i> , <a href="#">ACL 2020</a> [Paper] [Project Page]	

- Gupta, A., **Gupta, V.**, *Unsupervised Contextualized Document Representation*, [SustainNLP 2021](#) at [EMNLP 2021](#) [Paper]
- Gupta, V.**, Saw A., Nokhiz, P., Gupta, H., Talukdar, P., *Improving Document Classification using Multi-Sense Embeddings*, [NAACL-SRW 2019](#) (Oral) & [ECAI 2020](#) (Oral) [Paper] [Blog]
- Gupta, V.**, Bharti P., Nokhiz, P., Karnick, H., *SumPubMed: Summarization Dataset of PubMed Scientific Articles*, [ACL-IJCNLP SRW 2021](#) [Preprint] [Dataset]
- Gupta, V.**, Kumar, A., Nokhiz, P., Netrapalli, P., Rai, P., Talulkdar, P., *P-SIF: Document Embeddings using Partition Averaging*, [AAAI 2020](#) (Oral) [Paper] [Appendix] [PPT] [Poster] [Blog]
- Gupta, V.**, Nokhiz, P., Dutta, C., Venkatasubramanian, S., *Equalizing Recourse Across Group*, ArXiv 2019 [PrePrint]
- Uppal S., **Gupta, V.**, Swaminathan A., Zhang H., Mahata D., Gosangi R., Shah. R., Stent A., *Two-Stage Classification using Recasted Data for Low Resource Settings*, [AAACL-IJCNLP 2020](#) [Paper]
- Raunak, V., Dalmia, S., **Gupta, V.**, Metze, F., *On Long-Tailed Phenomena in Neural Machine Translation*, [EMNLP 2020](#) (Findings) & [SPNLP 2020](#) [Paper]
- Yadav, P., Yadav, P., Nokhiz, P., **Gupta, V.**, *Unbiasing Review Ratings with Tendency based Collaborative Filtering*, [AAACL-IJCNLP SRW 2020](#) [Paper]
- Li, T., **Gupta, V.**, Mehta, M., Srikumar, V., *A logic-Driven Framework for Consistency of Neural Models*, [EMNLP-IJCNLP 2019](#) & [StarAI 2020](#) [Paper] [Poster]
- Gupta, V.**, Wadbude, R., Natarajan, N., Karnick, H., Jain, P., Rai, P., *Distributional Semantics meet Multi-Label Learning*, [AAAI 2019](#) (Oral) [Paper] [Slides] [Poster]
- Gupta, V.**, Saw A., Gupta, H., Nokhiz, P., Talukdar, P., *Word Polysemy Aware Document Vector Estimation*, [NAACL-SRW 2019](#) (non-archival) [PrePrint]
- Raunak, V., **Gupta, V.**, Metze, F., *Effective Dimensionality Reduction for Word Embeddings*, [RepL4NLP 2019](#) (Oral) [Paper] [Poster]
- Raunak, V., Kumar, V., **Gupta, V.**, Metze, F., *On Dimensional Linguistic Properties of the Word Embedding Space*, [ACL-SRW 2019](#) (non-archival) & [RepL4NLP 2020](#) (Oral) [Paper]
- Dohare, S., **Gupta, V.**, Karnick, H., *Unsupervised Semantic Abstractive Summarization*, [ACL-SRW 2018](#) [Paper] [Poster]
- Mekala, D., **Gupta, V.**, Paranjape, B., Karnick, H., *Sparse Composite Document Vectors using soft clustering over distributional representations*, [EMNLP 2017](#) (Oral) [Paper] [Slides]
- Gupta, V.**, Karnick, H., Bansal, A., Jhala, P., *Product Classification in E-Commerce using Distributional Semantics*, [COLING 2016](#) (Master Thesis Work) [Paper] [Poster]
- Wadbude, R., **Gupta, V.**, Mekala, D., Karnick, H., *User Bias Removal in Review Score Prediction*, [CODS-COMAD 2018](#) & [DAB 2017](#) (Oral) [Paper] [Poster]
- Gupta, V.**, Mittal, S., Bhaumik, S., Roy, R., *Assisting Humans to Achieve Optimal Sleep by Changing Ambient Temperature*, [BIBM 2016](#), [BHI 2016](#) & [HI-DS 2016](#) (Oral) [Paper] [Slides]
- Mekala, Dheeraj., **Gupta, V.**, Kar, P., Karnick, H., *Bayes-optimal Hierarchical Classification over Asymmetric Tree-Distance Loss*, Tech Report [PrePrint]
- Mahajan, D., **Gupta, V.**, Keerthi, S., Sundararajan, S., *Efficient Estimation of Generalization Error and Bias-Variance Components of Ensembles*, Tech Report [PrePrint]

MASTER  
THESIS

### Product Categorization in E-Commerce using Distributional Semantics

Prof. Harish Karnick (IIT Kanpur) & Pradhuman Jhala (Flipkart.com)

[Thesis] [Slides]

- Proposed a novel distributional semantics representation for text documents.
- Use the proposed representation with a novel two-level ensemble approach utilizing (with respect to the taxonomy tree) a path-wise, node-wise and depth-wise classifiers for product classification.

RESEARCH  
INTERNSHIPS

Bloomberg AI (IE and KG), New York (Part-Time)

Aug 2021 - Dec 2021

---

Information Extraction for Trustworthy Tabular Reasoning <i>Vempala, Dr. Yujie He, Dr. Temma Choi</i>	<i>Dr. Shuo Zhang, Dr. Alakananda</i>
<b>Bloomberg AI (IE and KG), New York</b>	May 2021 - Aug 2021
Information Extraction for Trustworthy Tabular Reasoning <i>Vempala, Dr. Yujie He, Dr. Temma Choi</i>	<i>Dr. Shuo Zhang, Dr. Alakananda</i>
<b>Facebook AI Research (Assistant), Seattle</b>	Sep 2020 - Dec 2020
Efficient and Effective Semantic Parsing	<i>Dr. Denis Savenkov (Research Scientist)</i>
<b>IBM Research, Thomas J. Watson Research Center</b>	May 2019 - Aug 2019
Contrastive Explanations for Natural Language Task	<i>Dr. Kush R Varshney (Research Manager)</i>
<b>Microsoft Research India, Bangalore</b>	May 2016 - Jul 2016
Estimation of Generalization Error for Ensembles <i>Applied Scientist</i>	<i>Dr. Sundararajan Sellamanickam (Principal)</i>
<b>Flipkart Internet Pvt. Ltd., Bangalore (Part Time)</b>	Aug 2015 - July 2016
Web Scale Product Classification	<i>Pradhuman Jhala (Principal Architect)</i>
<b>Flipkart Internet Pvt. Ltd., Bangalore</b>	May 2015 - July 2015
Web Scale Product Classification	<i>Pradhuman Jhala (Principal Architect)</i>
<b>Samsung R&amp;D Institute, Bangalore</b>	May 2014 - July 2014
Mobile and Healthcare Solution Y2014	<i>Sandip Bhaumik (Group Manager) &amp; Raj Roy (Manager)</i>
<b>Synopsys Inc., Bangalore</b>	May 2013 - July 2013
DALI Verification System Coverage Visualization	<i>Yogesh Pandey (Group Director)</i>

## Teaching Experience

**Guest Instructor:** for CS 2420 - Introduction to Algorithms and Data Structures, Spring 2022. Taught two lecture on Graph Data Structures and Algorithms.

**Guest Instructor:** for CS 2420 - Introduction to Algorithms and Data Structures, Fall 2021. Taught two lecture on Graph Data Structures and Algorithms. Design the weekly assignment and the practice lab session.

**Seminar Coordinator:** for CS 7941-Data Science Seminar (Spring 2021, Fall 2021-2020, Summer 2020) along with two University of Utah professors (Prof. Jeff Phillips and Prof. Aditya Bhaskara)

**Teaching Mentor:** for CS 6355 Spring 2021 Structured Prediction. Involves office hours for doubt clearing, assignment and examination evaluation.

**Teaching Mentor:** for CS 6150 Fall 2019 Advanced Algorithms. Involves office hours for doubt clearing, assignment and examination evaluation.

**Teaching Assistant:** for MLT 2016 - Machine Learning Tool and Technique: Mentored a group of 30 M-Tech students part of a course on Machine Learning. Set up a labeling software for project work, resulting in a new dataset.

**Teaching Assistant:** for OS 2016 - Operating System: Mentored a group of 30 M-Tech students part of an introductory course on Operating System.

MISCELLANEOUS [\[Professional Services\]](#), [\[Seminars Talks\]](#), [\[Research Grants/Awards\]](#), [\[Mentored Students\]](#)