Shruti Singh

📞 9537160842
✉ singh_shruti@iitgn.ac.in
🖹 shruti-singh.github.io
in singh-shruti
⬤ shruti-singh

Education

**Aug'19 –
exp.
Dec'24**
Ph.D. Computer Science, GPA: 9.2/10.
Indian Institute of Technology Gandhinagar, India.
Advisor: Mayank Singh

**Aug'13 –
May'17**
BTech. Information and Communication Technology minors in Computational Science, GPA: 8.25/10.
DA-IICT, Gandhinagar, India.

Research Expertise

Natural Language Processing: *Introduced novel framework for adapting LMs to long documents using a contrastive loss with margin gradation [3], training, fine-tuning, and instruction-tuning LMs [2, 3, 11, 12], QA pipelines, datasets and benchmarks for in-depth understanding of scientific text [1, 2, 5, 6, 12], RAG [1, 2, 12], LoRA [12], training small LLMs.*

Data Science: *Developed a portal for recommending research papers and research-releated tweets, integrating advanced analytics to support researchers in knowledge exploration and discovery [7, 10].*

Computational Neuroscience: *Experience working with EEG data [4], ML models for meditation data [4], inspection of bilingual mental lexicon for bilingual speakers [13].*

Research and Industry Experience

**Aug'23 –
Jan'24**
YaleNLP Lab, USA
Fulbright-Nehru Visiting Researcher at Yale University, under supervision of Prof. Arman Cohan.

- Developed a novel reading comprehension dataset [1] that challenges language models to understand scientific articles deeply. The dataset necessitates reasoning across figures, tables, equations, appendices, and supplementary materials and introduces multi-document reasoning.
- Part of the team that curated a dataset of 137k instruction-following demonstrations for 54 tasks covering five scientific literature understanding capabilities [11].

**Aug'22 –
Dec'22**
Bosch Research, India
Research Intern, under supervision of Rishabh Gupta.

Worked on representing procedural documents as a concept graph. Proposed and implemented two algorithms to sparsify joint concept interaction graphs and incorporate sequential information in documents [14].

**Feb'18 –
June'19**
Enago Read, India
Research Engineer, under supervision of Dr. Sourish Dasgupta.

Accomplished a 147% increase in recommendation click rate by developing RAx Bot, a research paper recommendation agent that adapts to researchers' needs based on their resource selection patterns, for an AI-powered research assistant platform.

Jan'17 – **Sprinklr Inc., India**
Feb'18 *Product Engineer*.
- Worked on near de-duplication of large-scale real-time incoming web data of the scale of 25 million documents/day, ensuring minimal latency.
- Setup the natural language pipeline for Asia Pacific languages on the Sprinklr Listening platform. Implemented a language detection module for close languages Malay and Indonesian.

Publications

[1] **Shruti Singh**, Nandan Sarkar, and Arman Cohan, *"SciDQA: A Deep Reading Comprehension Dataset over Scientific Papers"* EMNLP 2024.

[2] **Shruti Singh**, Shoaib Alam, Husain Malwat, and Mayank Singh, *"LEGOBench: Scientific Leaderboard Generation Benchmark"* EMNLP Findings 2024.

[3] **Shruti Singh** and Mayank Singh, *"CoSAEmb: Contrastive Section-aware Aspect Embeddings for Scientific Articles"* Scholarly Document Processing Workshop at ACL 2024.

[4] **Shruti Singh**, Pankaj Pandey, Shivam Chaudhary, Krishna Miyapuram, and Derek Lomas, *"Towards Development of Personalized and Generalized Interfaces for Brain Signals across Different Styles of Meditation"* ICGVIP 2022.

[5] **Shruti Singh** and Mayank Singh, *"The Inefficiency of Language Models in Scholarly Retrieval: An Experimental Walk-through"* ACL Findings 2022.

[6] **Shruti Singh**, Mayank Singh, and Pawan Goyal, *"COMPARE: A Taxonomy and Dataset of Comparison Discussions in Peer Reviews"* Joint Conference on Digital Libraries 2021.

[7] Viraj Shah, **Shruti Singh**, and Mayank Singh, *"TweeNLP: A Twitter Exploration Portal for Natural Language Processing"* Demonstrations, ACL-IJCNLP 2021.

[8] Shriraj Sawant*, and **Shruti Singh***, *"Understanding Attention: In Minds and Machines"* ML-Retrospectives, Surveys & Meta-Analyses Workshop at NeurIPS 2020.

[9] Mihir Jain*, Tanmay Sharma*, Aman*, **Shruti Singh**, and Mayank Singh, *"Genealogy Tree Construction of Research Papers"* Extended Abstract, Young Researcher Symposium, CODS-COMAD 2021.

[10] Monarch Parmar*, Naman Jain*, Pranjali Jain*, Jayakrishna Sahit*, Soham Pachpande*, **Shruti Singh**, and Mayank Singh, *"NLPExplorer: Exploring the Universe of NLP Papers"* Demonstrations, European Conference on Information Retrieval 2020.

Under Review / Working Manuscripts.

[11] David Wadden, Kejian Shi, Jacob Morrison, Aakanksha Naik, **Shruti Singh**, Nitzan Barzilay, Kyle Lo, Tom Hop, Luca Soldaini, Shannon Zejiang Shen, Doug Downey, Hannaneh Hajishirzi, and Arman Cohan, *"SciRIFF: A Resource to Enhance Language Model Instruction-Following over Scientific Literature"* Under submission at NeurIPS Datasets and Benchmarks 2024.

[12] **Shruti Singh**, Husain Malwat, Xhitij Choudhary, Hitesh Lodwal, Rakesh Thakur, and Mayank Singh, *"Unlocking LLM Insights: A Dataset for Automatic Model Card Generation"* Under submission at CIKM 2024.

[13] **Shruti Singh** and Jooyoung Kim, *"Understanding Cross-Language Search in the Mental Lexicon: Semantic priming of Hindi and English"* Working manuscript.

[14] **Shruti Singh** and Rishabh Gupta, *"Sparse Graph Representations for Procedural Instructional Documents"* arXiv preprint.

Awards

Year	
2023	Awarded Fulbright-Nehru Doctoral Research Fellowship to Yale University.
2022	Selected for the ORKG Curation Grant 2022 by TIB Hannover, Germany.
2021	Awarded the Prime Minister Research Fellowship (PMRF).
2020	Selected for DLRL Summer School 2020 hosted by CIFAR and Mila, among 250 out of 1000+ applicants.
2013	Certificate of Merit for being in top 1.5% students in the CBSE AISSCE 2013.
2011	Dhirubhai Ambani SSC Merit Reward for outstanding performance at CBSE AISSCE 2011.

Teaching

Deep Learning. Graduate Teaching Fellow. IIT Gandhinagar (Spring 2023)

Natural Language Processing. Teaching Assistant. DAIICT (Spring 2023), IIT Gandhinagar (Autumn 2020)

Recommendation Systems. Teaching Assistant. DAIICT (Spring 2022)

Introduction to Data Mining. Teaching Assistant. DAIICT (Autumn 2022)

Databases. Teaching Assistant. IIT Gandhinagar (Spring 2023, Spring 2022, Spring 2021, Spring 2020)

Introduction to Python. Teaching Assistant. IIT Gandhinagar (Autumn 2021, Autumn 2019)

Professional Service

Program Committee Member for CIKM 2023 and CIKM 2024, Reviewer for ICLR LLM Agents Workshop 2024.

Delivered a talk and hands-on session on Autonomous Agents at ACM India Summer School on Generative AI for Text 2024.

Student organizer for ACM India Summer School on Generative AI for Text 2024.

Student Volunteer for ACM-IKDD Summer School 2022 on Data Science, NAACL 2021, ACL-IJCNLP 2021, IndoML 2020, ACL 2020, ICLR 2020, ACM-W India Workshop 2020, ACM India Annual Event 2020, and the Inter-Research-Institute Student Seminar (IRISS) 2020.

Webmaster for IndoML 2020

Technical Skills

Programming. Proficient in Python. Significant experience in C, C++, Java, Bash, HTML/CSS, SQL.

Tools and Technologies. Elasticsearch, MongoDB, Flask, Docker, Git, Slurm, CUDA, OpenMP, StanfordCoreNLP, Spacy, Wordnet, NLTK, OpenCV, Gensim, Sklearn, Surprise, PyTorch, LlamaIndex, LangChain, Apache Kafka, Apache Lucene.