Shruti Singh

(+1) 650-408-0884 | shrutis1051@gmail.com | shrutisingh@microsoft.com | shruti-singh.github.io
shruti-singh | singh-shruti | Google Scholar

Research Engineer at Microsoft Research specializing in foundation models, agentic workflows, and information retrieval. PhD from IIT focused on scientific reasoning models, supported by the Prime Minister's Research Fellowship (PMRF) and Fulbright Fellowship at Yale University. Strong publishing record in top-tier A* conferences (EMNLP, ACL).

Experience

Research Engineer, Microsoft Research *Apr 2025 - Present*

- Developed foundation models for information retrieval as part of the ReFoRM team led by Dr. Manik Varma.
- Scaled synthetic data generation for enterprise-specific domains.
- Spearheading efficient tokenizer and vocabulary construction strategies to optimize inference performance.

Senior ML Engineer, Qlik *Dec 2024 - Mar 2025*

- Developed a multi-agent orchestration system to automate end-to-end enterprise data pipelines, driving autonomous ingestion, complex analysis, and dynamic report generation.

Fulbright-Nehru Visiting Researcher, Yale University *Aug 2023 - Jan 2024*

- Developed SciDQA, a deep reading comprehension dataset requiring reasoning across figures, tables, equations, appendices, and supplementary materials.
- Co-developed SciRIFF with 137k instruction-following examples spanning 54 scientific literature understanding tasks.

PhD Research Intern, Bosch Research *Aug 2022 - Dec 2022*

- Architected graph representations for procedural documents and proposed algorithms to sparsify joint concept interaction graphs with sequential information.

Research Engineer, Enago Read *Feb 2018 - Jun 2019*

- Accomplished a 147% increase in recommendation click rate by developing RAx Bot, a research paper recommendation agent adapting to researchers' resource selection patterns.

Product Engineer, Sprinklr Inc. *Jan 2017 - Feb 2018*

- Worked on near de-duplication of large-scale real-time incoming web data at the scale of 25 million documents/day, ensuring minimal latency.
- Architected the natural language understanding pipeline for Asia Pacific languages, including language detection, tokenization, lemmatization, and keyword identification.

Education

Ph.D. in Computer Science *Aug 2019 - Jun 2025*
Indian Institute of Technology *Gandhinagar, India*
Thesis: Datasets, Models, and Tools for Scholarly Documents

B. Tech. (Hons.) in ICT with minor in Computational Science *Jul 2013 - May 2017*
Dhirubhai Ambani University *Gandhinagar, India*

Skills

Research Interests	AI for Science, model pretraining, efficient tokenizers, synthetic data, evaluation, agentic workflows, search and recommendation, RAG, LLM evaluation & benchmarking, leaderboard generation, enterprise data
Programming	Python, Bash, C++, Java, SQL
Tools & Technologies	PyTorch, LangChain, LangGraph, Flask, FastAPI, Docker, Git, Slurm, Kubernetes, Elasticsearch, MongoDB, Faiss, CUDA, OpenMP, scikit-learn, Surprise

Publications

- David Wadden, Kejian Shi, Jacob Morrison, Aakanksha Naik, Shruti Singh, et al., *SciRIFF: A Resource to Enhance Language Model Instruction-Following over Scientific Literature*. **EMNLP 2025**.

- Shruti Singh, Nandan Sarkar, and Arman Cohan, *SciDQA: A Deep Reading Comprehension Dataset over Scientific Papers*. **EMNLP 2024**.
- Shruti Singh, Shoaib Alam, Husain Malwat, and Mayank Singh, *LEGOBench: Scientific Leaderboard Generation Benchmark*. **EMNLP Findings 2024**.
- Shruti Singh and Mayank Singh, *CoSAEmb: Contrastive Section-aware Aspect Embeddings for Scientific Articles*. SDP Workshop at **ACL 2024**.
- 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**.
- Shruti Singh and Mayank Singh, *The Inefficiency of Language Models in Scholarly Retrieval: An Experimental Walkthrough*. **ACL Findings 2022**.
- Shruti Singh, Mayank Singh, and Pawan Goyal, *COMPARE: A Taxonomy and Dataset of Comparison Discussions in Peer Reviews*. **JCDL 2021**.
- Viraj Shah, Shruti Singh, and Mayank Singh, *TweeNLP: A Twitter Exploration Portal for Natural Language Processing*. **ACL-IJCNLP Demo 2021**.
- Shriraj Sawant and Shruti Singh, *Understanding Attention: In Minds and Machines*. ML-Retrospectives Workshop at **NeurIPS 2020**.
- Mihir Jain, Tanmay Sharma, Aman, Shruti Singh, and Mayank Singh, *Genealogy Tree Construction of Research Papers*. **CODS-COMAD YRS 2021**.
- Monarch Parmar, Naman Jain, Pranjali Jain, Jayakrishna Sahit, Soham Pachpande, Shruti Singh, and Mayank Singh, *NLPExplorer: Exploring the Universe of NLP Papers*. **ECIR Demo 2020**.
- Shruti Singh, Husain Malwat, Xhitij Choudhary, Hitesh Lodwal, Rakesh Thakur, and Mayank Singh, *Unlocking LLM Insights: A Dataset for Automatic Model Card Generation*. arXiv preprint.
- Shruti Singh and Rishabh Gupta, *Sparse Graph Representations for Procedural Instructional Documents*. arXiv preprint.

Teaching

Graduate Teaching Fellow & Teaching Assistant - IIT Gandhinagar
Deep Learning (Spring ’23), Databases (Spring ’20-’23), and Intro to Python (Autumn ’19, ’21). Designed coursework material, assignments, conducted hands-on tutorials, graded projects and mentored students.

Teaching Assistant - Dhirubhai Ambani University
NLP (Spring ’23, Autumn ’20), Recommendation Systems (Spring ’22), and Data Mining (Autumn ’22). Designed assignments, conducted hands-on tutorials and lab evaluations.

Honors & Awards

2023 **Fulbright-Nehru Doctoral Research Fellowship**, Yale University
2022 **ORKG Curation Grant**, TIB Hannover
2021 **Prime Minister Research Fellowship (PMRF)**, Government of India
2020 **Selected Participant**, DLRL Summer School hosted by CIFAR and Mila
2013 **Certificate of Merit**, Top 1.5% students in CBSE All India Senior Secondary Exam
2011 **Dhirubhai Ambani SSC Merit Reward**, Outstanding performance in CBSE All India Exam

Professional Service

- **Reviewer / Program Committee**: ARR (ACL Rolling Review), CIKM 2023-2026, ICLR LLM Agents Workshop 2024
- **Speaker & Session Lead**: Talk and hands-on session on Autonomous Agents at ACM India Summer School on Generative AI for Text 2024
- **Student Organizer**: ACM India Summer School on Generative AI for Text 2024
- **Student Volunteer**: NAACL 2021, ACL-IJCNLP 2021, IndoML 2020, ACL 2020, ICLR 2020, ACM-W India Workshop 2020, ACM India Annual Event 2020, IRISS 2020
- **Webmaster**: IndoML 2020