

Shrey Pandit

shrey pandit@utexas.edu | [linkedin.com/in/shrey pandit/](https://www.linkedin.com/in/shrey pandit/) | github.com/ShreyPandit | Austin, TX | +15123006621

EDUCATION

The University of Texas at Austin, Austin, TX

May 2023 – Present

*Masters in **Computer Science***

Relevant Coursework: Topics in Natural Language Processing, Reinforcement Learning, Database Management, Statistical Methods, Speech Processing

Birla Institute of Technology and Science, Pilani, IN

May 2019 – June 2019

*Bachelor of Engineering, Major: **Computer Science**, Minor: **Data Science***

Relevant Coursework: Machine Learning, Database management, Data Structures and Algorithms, Data Science

EXPERIENCE

Salesforce Research, Palo Alto, California

June 2024 – August 2024

AI Research Intern

- Created a novel modification in LLM structure that results in competitive performance in RAG tasks, surpassing GPT-4 and Command R Plus models while being 10x-100x smaller in size.
- Curated high quality Faithful evaluation dataset, that checks faithfulness and hallucination problems in LLMs.

TAUR Lab, University of Texas at Austin, TX

July 2023 – June 2024

Graduate Research Assistant

- Pioneered efficient debugging methods for correcting LLM-generated programs, employing feedback and error trace techniques tailored to LLMs. Achieved significant improvements in model accuracy & error rates.

Microsoft Research, Bengaluru, India

July 2022 – Jan 2023

Research Intern

- Spearheaded the compression of large language models using adapters, achieving a balance between computational efficiency & bias. Successfully deployed models, enhancing speed and fairness in applications.

Google Summer of Code, Online

June 2022 – September 2022

Student Collaborator

- Engineered a novel end-to-end multimodal vision transformer to detect hand gestures in TV news clippings, facilitating the creation of accessible captions for the specially-abled.

Princeton NLP Lab, Online

June 2021 – May 2022

Informal Research Collaborator

- Innovated a unique data augmentation method to bolster the performance of low-resourced languages in transformer models. This approach led to improvements in language processing capabilities.

PUBLICATIONS

- **FaithEval: Can Your Language Model Stay Faithful to Context** - ICLR 2025
- **SFR-RAG: Towards Contextually Faithful LLMs** - [Technical Report](#), 2024
- **CodeUpdateArena: Benchmarking Knowledge Editing on API Updates** - Preprint, Under Review, 2024
- **AdaPT: A Set of Guidelines for Hyperbolic Multimodal Multilingual NLP** - NAACL, 2024
- **A Comparative Study of Model Compression Techniques on Fairness in Language Models** - ACL, 2023
- **Can LLMs solve generative visual analogies?** Interactions between Analogical Reasoning & ML @IJCAI, 2023
- **DMIX: Adaptive distance-aware interpolative mixup** ACL, 2022
- **CIAug: Equipping Interpolative Augmentation with Curriculum Learning** NAACL, 2022
- **An Autoencoder-Based Approach to Simulate Sports Games** -MLSA@PKDD/ECML, 2020

TECHNICAL SKILLS

- Proficient: Python, C++, C, R, SQL, PyTorch, Tensorflow, Data Analysis
- Familiar: HTML, Git, Java, AWS, Rest API, Docker, Flask, Kubernetes, React, Django

AWARDS

- **Winner of Turing's Large Scale Models for Inclusion Hackathon Challenge, 2022** by Microsoft Turing- Implemented Inclusivity Toolkit to diagnose the biases of language models across various dimensions.
- **Google Research Week 2023** Among the top 250 students selected across India to attend a sponsored seminar on ML. Offered seat in the NLP track (among 50 students) in line with previous research experience.