

DIVIJ BAJAJ

divij.bajaj@tamu.edu | <http://linkedin.com/in/divij-bajaj-91292482/> | 9797212535 | Bryan, TX

Summary

Experienced professional with over 1.5 years of expertise in Machine Learning Research spanning **Computer Vision and NLP** and 2 years of experience in **Data Engineering**. Approaches tasks with a strong sense of responsibility and a genuine passion for continual improvement.

Education

Texas A&M University, College Station, TX

Texas, United States

Master of Computer Science, **GPA: 4.0/4.0**

Aug 2023 - May 2025

Relevant Coursework: Algorithms, Machine Learning, Deep Learning, Distributed Processing Systems

Indian Institute of Technology (BHU), Varanasi

Varanasi, India

Bachelor of Technology in Electronics Engineering

July 2017 - May 2021

Relevant Coursework: Computer Vision, Natural Language Processing, Linear Algebra, Statistics & Probability

Technical Skills

Languages: C, C++, Java, Groovy, Python, HTML, CSS, JavaScript, SQL

Others: TensorFlow, Keras, PyTorch, OpenCV, Pandas, Docker, SciPy, NumPy, Linux, Scikit-Learn, Microsoft Office, Microsoft Azure

Experience

NLP Lab, Dept of Computer Science, Texas A&M University

College Station, TX

Graduate Research Assistant (Advised by Dr. Ruihong Huang)

Oct 2023 - Present

- Curated a novel dataset containing parallel short stories with a male and female protagonist performing moral/immoral actions.
- Evaluated current closed and open-source Large Language Models (LLMs) to identify the presence of gender bias in making morality judgments using the curated dataset.

Wells Fargo

Bangalore, India

Data Engineer (Machine Learning)

July 2021 - Aug 2023

- Enhanced Capital Markets Data Ingestion Pipeline to fetch and process **5 million trading records** daily from various Order Management Systems and prepare them for the **Data Analytics ML model**.
- Built **order execution hierarchy** for **1 million** daily trading records using Apache Kafka for smoother data visualization.

International Institute of Information Technology, Hyderabad

Hyderabad, India

Undergraduate Research Assistant

Feb 2021 - Aug 2021

- Developed a novel smoothed Bayesian sample stratification approach in **PyTorch** to enable balanced mini-batch sampling in large, heavy-tailed Computer Vision datasets, achieving a **25% reduction** in standard deviation.
- Executed a novel data augmentation approach to reduce skewness in existing datasets by **over 40%**.

Blue Sky Analytics

Gurugram, India

Data Analyst Intern

May 2019 - July 2019

- Performed **web crawling** to collect hourly pollution data for the past 10 years from various national and international government websites, accruing more than **200,000 data points**.
- Created visualizations on temperature and rainfall trends highlighting the adverse effects of climate change.

Publications

- [1] Divij Bajaj, Yuanyuan Lei, Jonathan Tong, Ruihong Huang (2024). Evaluating Gender Bias of LLMs in Making Morality Judgements. Conference on **Empirical Methods in Natural Language Processing (EMNLP) (Findings)**
- [2] Shivapuja, S. V., Khamkar, M. P., Bajaj, D., Ramakrishnan, G., & Sarvadevabhatla, R. K. (2021). Wisdom of (Binned) Crowds: A Bayesian Stratification Paradigm for Crowd Counting. In Proceedings of the **29th ACM International Conference on Multimedia (MM '21)**. Association for Computing Machinery, New York, NY, USA, 3574–3582.
<https://doi.org/10.1145/3474085.3475522>

Projects

Selecting Optimal Context from Long Documents in Domain-Agnostic Manner

Supervisor: Prof. (Dr.) Ruihong Huang, Natural Language Processing Lab

Nov 2023 - Apr 2024

- Designed an approach to retrieve the sentences relevant to a target sentence from long documents that do not fit in the limited input size of current **Large Language Models (LLMs)**.
- Evaluated the method on various downstream NLP tasks like **information retrieval and question answering**.