

SATYA SAURABH MISHRA

Gangeshwar Nath Chunar, 231304, Mirzapur U.P

+91 6388551787

✉ samrewa19@gmail.com

🌐 saurabh-mishra-049424177

🔗 satyam19mishra

Education

PDPM, IIIT Jabalpur

M.Tech - Artificial Intelligence and Machine Learning — CGPA: 8.8

Jabalpur, MP

Aug. 2022 - Aug. 2024

Experience -

Data Scientist — Dell Technologies (July 2024 - Present), Bengaluru

ICLR 2025 Accepted Research Paper: How Does Entropy Influence Modern Text-to-SQL Systems?

- Proposed an **entropy-based metric** to assess and **cluster SQL query candidates for improved confidence estimation**.
- Integrated entropy-guided refinement into **CHESS** and **CHASE** pipelines to reduce redundant generations.
- Built a modular agentic framework using **Mixtral** and **Deepseek** models for **retrieval, generation, and selection**.
- Applied **MarkupLM** with **DBSCAN** to cluster execution-based embeddings of SQL outputs.
- Used entropy thresholds as **stopping criteria in query refinement** to cut down **computation costs**.
- Ran detailed experiments on **BIRD** benchmark using diverse generation strategies like **Divide-and-Conquer, Query Plan, Online Synthetic Examples**.

AI-Powered Balance Sheet Flux Commentary Automation

- Automated **6,000 quarterly flux analyses** using a custom-built LLM pipeline, **reducing 32K+ hours** of manual effort annually.
- Conducted extensive prompt engineering experiments and adopted **Chain-of-Thought** prompting for best results.
- Fine-tuned a **Llama-3.1-8B-Instruct** using **PEFT** with **LORA** on historical finance data to improve accuracy.
- Designed and implemented an **tool based agentic framework** to orchestrate data preprocessing, variance analysis, and auto-commentary generation.

Data Science Intern — Dell Technologies (July 2023 - May 2024), Bengaluru

Intelligence Search

- Enhanced **LLM model's query response using RAG methodology** with Dell data.
- Conducted end-to-end experiments on **llama-index** and **langchain**, integrating **OpenAI** and **custom LLM** models.
- Led efforts to **fine-tune embedding models**, improving their ability for enhanced dell data representation.
- Explored **Vector Database** like **PGVector, Chroma DB** and **Faiss** for efficient storage and **retrieval**.
- Implemented Different prompt techniques like **Chain of Thought** for optimizing and getting better results from LLM.
- optimized model performance** through different **evaluations strategy** and comparisons with other llm models.

OpenSource And Publications

- 1 - Open-sourced finance chat and embedding models finetuned on 50M-token financial datasets. 🧡
- 2 - ICLR25 WorkShop - How Does Entropy Influence Modern Text-to-SQL Systems? (**Publication pending.**)

Projects

E2E NLP Text Summarization With Deployment 🧡 | LLM, NLP, Docker, EC2, CI/CD Sep. 2023 - Oct. 2023

- Architected and deployed an E2E Text Summarization ML app with google pegasus model on AWS EC2, orchestrated CI/CD using github actions, and containerized with Docker for scalable deployment.

Amazon Reviewer Network Analysis 🧡 | Python, NetworkX, Matplotlib, Pandas, Numpy Oct. 2022 - Dec. 2022

- Analyzed and **identified-cohesive reviewer communities** within Amazon product reviewers. Uncovered valuable patterns in shared item reviews, enabling personalized **product recommendations** to community members.

Technical Skills

Languages and Tools : Python, C/C++, Git/GitHub, DagsHub, DVC, MLflow and Airflow.

Generative AI: **GenAI, LLM**, Sklearn, Keras, CNN, RNN, LSTM, NLTK, **Transformer, Prompt Technique, RAG**, TensorFlow, Networkx, Matplotlib, Pandas, Numpy, Pytorch.

Achievements And Position of Responsibility

- Received the **Game Changer Award** from CFO for impactful contributions to AI-driven solutions.
- Delivered a talk on **Advanced Prompt Engineering Techniques and Retrieval-Augmented Generation (RAG)** at an internal AI knowledge-sharing session.
- Qualified GATE in **2021 and 2022**.