MD SARFRAZ ALAM

https://mdsarfraz.tech/

📍 Bengaluru   📞 +91-7760449876   ✉ mdsarfraz0202@gmail.com   in mdsarfraz2002   ○ mdsarfraz2002

Experience

Krutrim   **Jul 2024 – Present**
Software Developer I   *Bengaluru*

- Annotation Tool Development: Built an internal tool for large-scale audio, image, and video annotation to generate synthetic data for TTS and ASR model training. Utilized Django, PostgreSQL, and Kubernetes for managing AWS instances and load balancing.
- Data Pipeline Creation: Designed a data pipeline using PySpark to process text, image, and video data from Common Crawl. Implemented deduplication, language identification (LangID), and data transformation.
- LLM Guardrail Development: Engineered guardrails for large language models (LLMs) to enable safe and secure interactions, applicable to any model.
- Indic Data Collection: Developed a pipeline for scraping Indic image and newspaper data from various sources, ensuring daily updates and data freshness.
- High-Performance Data Download Tool: Developed a full-stack internal tool using React and Flask to download terabytes of data from Hugging Face, Internet Archive, and similar sources within minutes. Leveraged multiprocessing, multithreading, and load balancing across multiple AWS instances for optimized performance.

Krutrim   **Feb 2024 – Jul 2024**
Software Developer Intern   *Bengaluru*

- Streamlined workflows & enhanced operational efficiency by developing internal tools with React, Next.js & Flask.
- Designed & implemented an annotation app using Flask & React, integrating REST APIs for seamless data communication.
- Ensured scalability & performance by utilizing MongoDB for efficient data storage & retrieval, while optimizing data querying processes with SQL. Conducted thorough testing & debugging using tools like Postman to ensure application reliability & accuracy.

OLA   **Aug 2023 – Nov 2023**
AI Prompt Intern   *Bengaluru*

- Conducted experiments to evaluate the performance of AI models, & Developed Pipelines for Annotation of models
- Fine-tuned models for industry-specific use cases, reducing bias & errors, and deploying large-scale AI models.
- Conducted experiments to evaluate the performance of AI models, and Developed Pipelines for Annotation of models

Technical Skills

Languages: Python, Java, JavaScript, SQL, MongoDB
Technologies/Frameworks: Spark, Kafka, Ansible, React, Flask, AWS, Node.js, Next.js, REST API, Django
Developer Tools: Git, VS Code, Firebase, Vercel, Docker, GitHub, Postman, Jira, New Relic
Other Skills: NLP, Generative AI, Web Scraping, Prompt Engineering, Fine-Tuning, Large Language Models (LLMs)

Education

Presidency University   **Aug. 2020 – Jun 2024**
Bachelor of Technology in Computer Science (Data Science)

Deeksha   **Jun. 2019 – Mar 2020**
II PU (PCMC)

Ryan International School   **Jun. 2017 – Mar 2018**
10th (ICSE)

Awards & Leadership

Analytics Attax   **Mar 2023**
Secured 3rd position in National level hackathon Organised By IIT Kanpur   *IIT Kanpur*

VisionTech Fest   **Dec 2023**
National Hackathon Winner   *Visionet Systems Inc.*

FORGE - Data Science Club   **Dec 2022– Jun 2024**
Core Team Lead   *Presidency University*

National Service Scheme(NSS)   **Jan 2021– Jun 2024**
Volunteer   *Presidency University*