



Rohan Mitra

Master Student

✉ mitrarm@gmail.com

☎ +49 15510253026

📍 Bahnhofstraße 55, Zi. 302, 94469 Deggendorf, Germany

About Me

Data engineer transitioning to AI, adept in ETL and databases, armed with expertise in large language models (LLMs) and NLP models, eager to leverage advanced analytics, AI, and machine learning expertise within new roles, pursuing a Master's in AI. Deeply curious about AI advancements and the future.

Date of birth: December 18, 1995, Place of birth: Kolkata, India

Work experience

1/2024 - 5/2024

Senior Data Engineer

Alternative Path • Remote

As a data Engineer for an investment-management firm client, I architect and optimize large-scale web scraping and ETL solutions to enhance data-driven decision-making. Leveraging advanced technologies, I focus on efficient data extraction, processing, and analysis.

Key aspects of my role include:

- Designing and implementing robust ETL pipelines using Scrapy, S3, Snowflake, and Apache Airflow for seamless data extraction and transformation.
- Utilizing AWS (S3, EC2, CloudWatch, Lambda) for scalable cloud infrastructure and managing data storage with Snowflake and PostgreSQL for high performance.
- Implementing monitoring and alerting systems with PagerDuty for quick issue resolution and minimal downtime.
- Coordinating with cross-functional teams using Jira for agile project management and continuous improvement.
- Leveraging additional tools like GitHub, Jenkins, Playwright, and Selenium for efficient development, testing, and deployment processes.

7/2019 - 3/2023

Software Engineer II

D. E. Shaw & Co. • Hyderabad, IN

As a Software Data Engineer, I developed, designed, and deployed ETL pipelines for the business. Leveraged core technologies - Python, PostgreSQL, S3, SQL Server, AWS to build high-quality data pipeline systems.

Used Airflow, Docker, Kubernetes, and Grafana for scalable production and observability. Built frameworks for metadata tracking, data governance, data monitoring, web scraping, and data extraction tools. Leveraged AWS, Apache Kafka, SQL Server, and Apache Airflow technologies.

Worked closely and added value as a go-to data person with users across Data scientists and business stakeholders.

Accomplishments include building high availability robust ETL data pipelines, and handling batch processing. Implemented data transformation tools like Spark, and Pandas to achieve high throughput.

7/2018 - 7/2019

Data Engineer

LatentView Analytics • Chennai, IN

Python Development with a focus on text data extraction, building and analyzing models with nltk and other machine learning libraries.

2/2018 - 7/2018

Intern

LatentView Analytics • Chennai, IN

Worked on building a machine learning model for matching the menu to the accurate recipe scraped across the web from different sources, using rule-based and SVM rank.

Education

2023 - present

Master

Technische Hochschule Deggendorf • Deggendorf, DE

Artificial Intelligence and Data Science

2014 - 2018

Bachelor

Kalinga Institute of Industrial Science • Bhubaneswar, IN

Computer Science and Engineering

Skills

Apache Airflow

Amazon Web Services (AWS)

Apache Kafka

ETL

GitHub

Jira

Linux

Natural Language Processing(NLP)

Pandas

PostgreSQL

PySpark

Python (Programming Language)

PyTorch

Test-Driven Development

Data engineering

Large Language Models

HuggingFace API

Amazon Redshift

Machine learning

Streamlit

Data Governance

Data visualisation

Kafka

SQL

OpenCV

Tensorflow

Docker

Languages

Englisch

Advanced

Deutsch

Beginner

Github and Website

<https://github.com/rohanmitro>

<https://rohanmitro.me/>

Research and Projects

An analysis and classification of Fake News Headlines using NLP Techniques

Link - <https://bit.ly/nlp-projekt>

Topic - Implemented a model for predicting fake news headlines using TF-IDF vectorization and other text analysis models.

Achievements

fast.ai Scholarship - Feb. 2018: Selected for a fast.ai cohort to attend online lectures.

MIT Startup Fellowship - Mar. 2016: Selected to present startup ideas among several participants.