

RUDRA PRASAD BHUYAN

AI/ML Engineer

Rourkela, Odisha | +91-6370874791 | rudraprasadbhuyan999@gmail.com | linkedin.com/in/rudraprasadbhuyan | github.com/Rudra-G-23

PROFILE SUMMARY

AI/ML Engineer experienced in production LLM systems, data pipelines, and client-facing AI products. Delivered an AI interview platform and an airline Talk-to-Data system, combining RAG, hybrid retrieval, Text-to-SQL, and evaluation workflows. Strong foundation in Python, SQL, data engineering, and translating complex data into deployable solutions.

TECHNICAL SKILLS

Programming: Python, SQL **ML & LLM:** Scikit-learn, PyTorch, TensorFlow, LLM fine-tuning, RAG, LangChain, LangGraph, LangSmith
Data: Pandas, NumPy, Polars, PySpark, PostgreSQL, BigQuery, Snowflake, Databricks, Star Schema **Tools:** Docker, AWS, GitHub Actions, MLflow, OpenTelemetry, Power BI, FastAPI, Flask

WORK EXPERIENCE

algen.ai

June 2026 - Oct 2026

AI/ML Engineer

- Delivered an AI interview platform end-to-end for a confidential client, covering LLM question generation, real-time audio processing, and cost-optimized inference.
- Designed 12 Star Schema databases that enabled up to 15 relevant, non-repetitive interview questions per request.
- Reduced resume-analysis cost to **INR 0.13 per student**, enabling analysis for approximately 100 students at about INR 13 total.
- Built an airline Talk-to-Data application using RAG to retrieve relevant tables, columns, and relationships from internal documentation.
- Implemented hybrid semantic and BM25 retrieval to ground a Text-to-SQL model in relevant business context.
- Designed a three-model workflow for question understanding, fine-tuned Text-to-SQL, and SQL-to-natural-language responses.
- Created a **400-sample** golden dataset to evaluate Talk-to-Data accuracy and reliability.
- Guided a **4-person** team on revenue-pipeline optimization for a **60-110 GB** aviation dataset.

SBC Labs

Nov 2025 - Feb 2026

Jr. Data Scientist

- Developed Python and SQL ETL pipelines for 15 multi-level HCES datasets with 400+ features.
- Analyzed 500+ socioeconomic and healthcare variables to identify actionable trends.
- Automated data validation and duplicate handling with SQL procedures, reducing inconsistencies by 20%.
- Built reusable preprocessing scripts, saving approximately one hour of manual preparation daily.
- Created 30+ analytical reports and KPI dashboards across 50+ parameters for stakeholder decision-making.

Tata Power WODL

June 2025 - July 2025

Quality Analyst

- Analyzed smart-meter data workflows for 60,000+ consumers and centralized database and issue-tracking processes.
- Supported field-to-database survey operations with a 20-member team, digital meter updates, and complaint handling.

SELECTED PROJECTS

Fine-Tuning a Code LLM for Python FIM Completion: Curated a Python-only dataset from The Stack v3 (133 TB) and fine-tuned Qwen2.5-Coder-0.5B; model and dataset published on Hugging Face.

Vehicle Insurance Risk Prediction: Built an MLOps pipeline with MongoDB, AWS S3/EC2, Docker, GitHub Actions, and a real-time Flask prediction app.

find-my-joint: Developed a Python package that profiles 400+ columns across 15-level datasets and recommends high-confidence joins with approximately 90% matching accuracy in under two minutes.

SQL Modern Data Warehouse: Built a PostgreSQL Medallion architecture, ERP/CRM ETL pipelines, star schema, and Power BI dashboards.

EDUCATION

Utkalmani Gopabandhu Institute of Engineering - Diploma, Electrical Engineering (2023 - 2026)

Govt. Autonomous College - Science (2020 - 2022)