

Dristanta Das

+91-8583918089 | dasdristanta13@gmail.com | github.com/dasdristanta13 | linkedin.com/in/dasdristanta13

Skills

AI/ML: NLP, LLMs, NER, RAG, Machine Learning, Deep Learning, Statistical & Predictive Modeling

Frameworks: LangChain, LangGraph, PyTorch, Hugging Face, Scikit-learn, XGBoost, FastAPI, Agentic AI

Tools & Cloud: Python, R, SQL, Git, Docker, BigQuery, Vector Databases, AWS, Azure

Design: Architecture Design, Data Flow Modeling, UML, Anomaly Detection, HIPAA/GDPR Compliance

Experience

Genpact

Bengaluru, Karnataka

Assistant Manager - Data Science

Apr 2025 – Present

- Architected and deployed an end-to-end **Text-to-SQL system** using **LangChain** and **LangGraph** for **50+ beta users**, supporting **10 concurrent users** and enabling natural language querying of complex databases, reducing query creation time by **65%**.
- Engineered an automated **invoice processing pipeline** leveraging **LLMs** for document parsing and information extraction, achieving **85%+ accuracy** across **10,000+ invoices** and reducing manual processing time by **60%**.
- Built robust **agentic workflows** with state management, error handling, and dynamic routing capabilities, achieving **92% task completion rate** and delivering superior user experience across LLM-powered applications.
- Spearheaded integration of LLM-powered solutions into production systems with cross-functional teams of **12+ members**, ensuring scalability to handle **500+ daily queries** and alignment with business requirements.

UST

Kolkata, West Bengal

Associate III Data Scientist

Oct 2022 – Mar 2025

- Orchestrated development of a client-facing provider search system using **advanced NLP techniques** and open-source **LLMs**, boosting search efficiency by **30%** and accelerating data preparation tasks by **40%**.
- Implemented **NER** and **semantic search** capabilities for retrieving medical terms from layperson language, achieving a **25%** improvement in operational efficiency and improving service quality for end-users.

Associate II Data Scientist

Jul 2022 – Sept 2022

- Conducted a patient data evaluation using **BigQuery** with comprehensive data transformations and statistical analysis, improving the client's understanding of adherence drivers by **45%** and improving outreach to more **than 10,000** patients.
- Designed and implemented a data anomaly detection system for categorical healthcare data, adhering to **HIPAA** and **GDPR** compliance standards.

Personal Projects

Retrieval-Augmented Generation (RAG) QA Bot Development

- Engineered a Python-based **QA bot** with hybrid retrieval using **BM25** and **DPR**, achieving **40%** higher answer accuracy compared to traditional search methods.
- Integrated **GPT-3.5-turbo**, **Phi-2**, and **Llama3** with contextual compression to reduce hallucinations by **60%**; optimized performance using **SQLite** caching and Slack integration.

Automated Summary Evaluation Competition

Organized by CommonLit, Vanderbilt University, Georgia State University & Kaggle

- Developed a machine learning model that gained a 0.48 Mean Columnar Root Mean Squared Error, ranking in the top **9%** of participants.
- Achieved a Bronze Medal on Kaggle by developing a solution for evaluating student-written summaries, enhancing evaluation accuracy by **65%** and outperforming **90%** of competing solutions.

Education

RKMVERI, Belur

July 2020 – Jun 2022

M.Sc. in Big Data Analytics

CGPA: 7.71/10.0

Presidency University, Kolkata

May 2017 – Jun 2020

B.Sc. in Mathematics

CGPA: 7.05/10.0

Certificate & Achievements

AWS Certified Cloud Practitioner: CLF C01

Microsoft Certified: Azure Fundamentals, AZ-900