



RUDRANI GHOSH

Data Scientist | Programmer | Lawyer
M: +91 8420870089 | Email: rudranighosh24@gmail.com
Portfolio Website: rudranig.com
<https://www.linkedin.com/in/rudrani-ghosh/>
<https://github.com/RudraniGhosh24>

EDUCATION

Master of Science in Data Science (M.Sc) **CGPA: 9.19** **July 2023 - May 2025**

Vellore Institute of Technology, Vellore

- Thesis on "Fine-Tuning LLM with RAG for Generating Educational Audio-Visual Content".
- Subjects: Statistical Inference, Linear Algebra, Probability, Machine Learning, Deep Learning, Natural Language Processing, Data Visualization, Time Series Analysis & Forecasting, Artificial Intelligence, Operations Research

Bachelor of Laws (L.L.B.) **CGPA: 9.7** **Aug 2020 - May 2023**

Techno India University, Kolkata

- Subjects: Constitutional Law, Criminal Law, Contract Law, Tort Law, Family Law, Property Law, Administrative Law, Jurisprudence, Evidence Law, Procedural Laws (CrPC & CPC), Intellectual Property Rights, Human Rights, Environmental Law, Labour and Industrial Law, International Law, Company Law, Alternative Dispute Resolution, Law of Arbitration and Conciliation.

Bachelor of Computer Applications (B.C.A.) **CGPA: 9.65** **Aug 2016 - May 2019**

Lovely Professional University, Punjab

- Subjects: Programming in C and C++, Java Programming, Data Structures, Database Management Systems (DBMS), Operating Systems, Computer Networks, Web Development (HTML, CSS, JavaScript), Software Engineering, Mobile Application Development, and Cybersecurity.

SUMMER SCHOOL EXPERIENCE

Vrije Universiteit Amsterdam (VU Amsterdam) **July 2024-August 2024**

Course: Logic as a Tool for Modelling | Grade: 9.0 | Credits: 3.0 | Credential ID: 2868745

Attended an advanced course focused on applying formal logic as a foundational tool for modeling complex real-world scenarios. Gained hands-on experience in logical reasoning, first-order and second-order logic, and their practical applications in AI, legal reasoning, and computational systems.

MASTER THESIS

Fine-Tuning LLM with RAG for Generating Educational Audio-Visual Content **April 2025**

This project introduces an AI-powered learning support system aimed at delivering accurate, grade-specific educational content aligned with CBSE and ICSE curricula for Classes 1 to 12.

WORK EXPERIENCE/ INTERSHIPS

Freelance AI Trainer - Outlier **Mar 2025 - Present**

- Working on training and evaluating machine learning and large language models across multiple domains. Responsibilities include assessing model outputs for accuracy, coherence, and domain-specific correctness in areas such as programming, law, and linguistics. Contribute to improving model performance through structured feedback, prompt refinement, and edge-case identification. Collaborated with distributed teams to enhance model reliability and alignment with real-world use cases.

Founder - LawReformer (<https://www.lawreformer.com/>) **Feb 2025 - Present**

- Built a legal-tech system [ai.lawreformer.com](https://www.lawreformer.com/) that includes multilingual, audio-supported AI legal assistance for underserved users. Also developed tools for legal research, clause analysis, statutory deadlines, and scenario simulation. Focused on access to justice, user-centered design, and applied AI reasoning.

Founder - Astroformer (<https://www.astroformer.com/>) **Jan 2025 - Present**

- Created and scaled an astrology-based platform offering automated insights on career growth, relationships, marriage timing, and other life aspects. Developed real-time interactive tools such as birth chart analysis and tarot readings, focusing on user engagement and high-CTR content strategies. Leveraged data analytics and AI-driven logic to personalize user experiences and improve retention.

Machine Learning Intern - Infosys Springboard **May 2024 - Jun 2024**

- Worked on unsupervised learning approaches for fraud pattern detection in healthcare datasets, including insurance fraud, identity theft, and phantom fraud scenarios.

Data Science Intern - IBM Skill Builders **Jan 2024**

- Explored machine learning and introductory NLP workflows through applied projects, including time-series forecasting and API integration using IBM Watson.

Machine Learning Intern - SystemTron **Feb 2024**

- Developed predictive models using regression and classification methods, including applications in healthcare, sports analytics, and price prediction.

ACHIEVEMENTS

Scholarship Recipient Vrije Universiteit Amsterdam (VU Amsterdam) - Summer School 2024

AWS AI ML Fellow 2023-2024 | **WiBD Shakti Fellow** 2024

University Gold Medal Recipient- Issued by Techno India University · Aug 2024

University Gold Medal Recipient- Issued by Lovely Professional University · Sep 2019

SKILLS

- **Programming:** Python (NumPy, Pandas, Scikit-learn, TensorFlow, Keras, NLTK, SpaCy, Matplotlib), R (ggplot2, dplyr, caret, tidyverse, randomForest), Prolog (SWI-Prolog), Git/GitHub, SQL
- **Machine Learning & AI:** Logistic Regression, Decision Trees, KNN, SVM, Random Forest, Deep Learning (ANNs, CNNs, RNNs), Computer Vision (OpenCV, Image Classification), Time Series Forecasting (ARIMA, LSTM)
- **NLP, LLMs & Knowledge Systems:** Natural Language Processing, Fine-Tuning LLMs (TinyLlama, Hugging Face Transformers), Retrieval-Augmented Generation (RAG), Prompt Engineering, Entity Extraction, Knowledge Representation, Logic Programming
- **Data Science & Optimization:** Feature Engineering, Data Preprocessing, Model Evaluation (Cross-Validation, GridSearchCV, Hyperparameter Tuning), Data Visualization (Matplotlib, Seaborn, Plotly), Multilingual Text and Speech Processing
- **Legal & Interdisciplinary Expertise:** Legal Reasoning, Jurisprudence, Statutory Interpretation, Legal Research & Drafting, Computational Legal Analysis

PROJECTS WORKED ON (UNPUBLISHED):

- **Classifying Offenses under the Indian Criminal Act**

Utilizes TF-IDF, GloVe embeddings, and WordNet to map user inputs to 358 sections of the Bhartiya Nyay Sanhita (2023). Provides legal details like punishment, cognizability, bailability, and jurisdiction, and auto-generates draft FIRs and bail bonds.

- **Automated Legal Clause Detection and Visualization**

Compares Bigram models and RNNs (with and without GloVe) to detect key clauses (e.g., payment terms, jurisdiction) in legal documents. Highlights detected clauses with bounding boxes to improve interpretability in legal review.

- **Bridging Legal Reasoning with Statistical NLP, RNNs, and First-Order Logic**

Integrates POS tagging, word embeddings, and First-Order Logic to model actus reus and mens rea. The hybrid framework enhances offense classification with both statistical accuracy and logical transparency.

- **DiagnoChat: A Multilingual, Multimodal Approach to Accessible Healthcare and Real-Time Disease Diagnosis**

A chatbot platform for disease prediction using text/image inputs. Employs ML models (SVM, KNN, RF, etc.) with synonym matching and a symptom dictionary. Supports six Indian languages and recommends local medical help.

- **State-Wise Analysis and Forecasting of Crimes Against Women in India**

Analyzes two decades of crime data with socio-economic features using PCA, Lasso, RFE, and DBSCAN. Reveals regional patterns and uses ARIMA to forecast crime trends, aiding in policy planning.

- **A Hybrid Approach to Legal Text Summarization and Entity Extraction Using Regex, POS Tagging, and Word Embeddings**

Extracts legal entities (acts, sections, parties, jurisdictions) using regex, POS, and embeddings. Summarizes judgments using a BART model and cosine similarity, balancing structure with clarity for legal use.

- **Fine-Tuning LLM with RAG for Generating Educational Audio-Visual Content**

Master's thesis project developing a multilingual learning platform using TinyLlama and GPT-4o Mini. Generates curriculum-aligned audio-visual content for classes 1-12 across six Indian languages, using RAG for real-time, accurate, inclusive education delivery.

CERTIFICATIONS

- **AWS Machine Learning Fundamental | Udacity | Issued Jul 2024**

<https://www.udacity.com/certificate/e/c7b6ccc6-f6c0-11ee-90a9-b3416383fa91>

- **Associate Data Scientist Track | DataCamp | Issued May 2024**

<https://www.datacamp.com/completed/statement-of-accomplishment/track/bb2282e3a5ecfc1d58122f0d17f0c078c45efa2f>

- **Data Scientist in Python Track | DataCamp | Issued May 2024**

<https://www.datacamp.com/completed/statement-of-accomplishment/track/ffacd58387c38374fef82633f413aac0ed0aeae2>

- **Enterprise Data Science in Practice | IBM | Issued Feb 2024**

https://www.credly.com/badges/e989017a-b387-4a4c-959d-70d788f3457d/public_url

- **Machine Learning for Data Science Projects | IBM | Issued Jan 2024**

https://www.credly.com/badges/849b383a-a25b-4eac-a42c-3d4779a9fb74/public_url