Ramnarayan Choudhary

NLP Engineer — Data Scientist — Software Engineer
Abu Dhabi, United Arab Emirates

• +971-581804812 • choudharyramnarayan123@gmail.com • LinkedIn • GitHub

Education

- **Mohamed bin Zayed University of Artificial Intelligence** Abu Dhabi, UAE
  Master of Science in Natural Language Processing *2025 – 2027**
  - Relevant Coursework: Deep Learning, Machine Learning, Advanced NLP, Speech Processing
- **Indian Institute of Technology (IIT) Bombay** Mumbai, India
  Bachelor of Technology in Chemical and Computer Science Engineering *2021 – 2025*
  - Relevant Coursework: Data Structures and Algorithms, Data Science, Entrepreneurship

Scholastic Achievements

- Awarded a fully funded scholarship by the UAE government
- Secured 99.31 percentile in JEE Main among 1.2M aspirants (2021).
- AIR 2305 in JEE Advanced among 250,000 aspirants (2021).
- AIR 667 in KVPY among 150,000 aspirants (2020).
- 396/450 in BITSAT, ranked among top 800 in India (2020).
- Recipient of prestigious Kishore Vaigyanik Protsahan Yojana (KVPY) Fellowship (2020).

Skills

- **NLP & LLMs:** Fairness & Safety, Privacy, Agentic Systems, Evaluation Methodology, Pre/Post-Training, Parameter-Efficient Optimization, fine Tuning, Robustness Analysis
- **Programming & Frameworks :** Python, C++, SQL, PyTorch, transformer, LangChain

Research Projects

- **Inference time Steering in Large Language Models** [GitHub]
  - Investigated inference-time activation steering as a unifying framework for controlling reasoning and alignment behaviors in LLMs, spanning formal reasoning and open-ended generation tasks.
  - Designed and evaluated contrastive steering methods that intervene directly in transformer hidden representations to modulate model behavior without parameter updates, enabling fine-grained control over plausibility bias, persona expression, and safety-related behaviors
- **PromptGCG: Gradient Coordinate Geometry for Prompt Sanitization** [GitHub]
  - Proposed a gradient-based prompt sanitization method that optimizes private token replacements to minimize privacy leakage while preserving downstream task utility in black-box LLMs.
  - Developed a formal privacy–utility evaluation framework with adversarial inference attacks, demonstrating that local surrogate models can effectively protect sensitive prompt content
- **Machine-Generated Code Detection** [GitHub]
  - Investigated machine-generated code detection under cross-language and multi-generator, focusing on representation learning challenges and generalization beyond generator-specific stylistic.
  - Analyzed how different adaptation regimes (full fine-tuning, PEFT, prompting) affect a model's ability to capture semantic versus superficial cues in code provenance classification.

Research and Professional Experience

- **Generative AI Engineer – AITHON** Mumbai, India
  May 2025 – Aug 2025
  - Designed an end-to-end, modular document intelligence system for ingestion, classification, and structured extraction using GPT-4o and Gemini 2.5 Pro, achieving 95%+ task accuracy.
  - Built scalable monitoring, including LangGraph-based intelligent routing for text/vision inputs and a real-time analytics dashboard with 15+ interactive modules.
- **AI Research Intern – MBZUAI** Abu Dhabi, UAE
  May 2024 – Jun 2024
  - Conducted research on LLM-based reasoning for Linguistic Olympiad problems
  - Developed multi-agent reasoning and benchmarking pipelines across GPT, LLaMA, and Mistral models, improving accuracy to 68% (+11% over strong baselines).
- **LLM R&D Intern – Attentions.ai** Pune, India
  Jul 2024 – Aug 2024
  - Fine-tuned long-context LLMs (Phi-3 128k) for structured SQL generation using PEFT methods, achieving 83% execution accuracy on enterprise queries.
  - Designed and deployed ML system on AWS and integrated it into a full-stack MERN application.
- **Computer Vision Intern – Assert AI Secure Tech** Mumbai, India
  May 2023 – Jun 2023
  - Developed a YOLO-based toll plaza surveillance system trained on 100K+ CCTV images.
  - Deployed the model on Jetson Nano using DeepStream, achieving 99.3% mAP under real-time.

Key Projects

- **New User Engagement Prediction** – Modeled next-month user activity from large-scale behavioral logs using ensemble learning; addressed severe class imbalance via resampling and boosting, achieving **49.56% F1** and **3rd place** among 850+ teams on the Zindi challenge.
- **Language Translation using Transformers** – Implemented Transformer architecture with self-attention, reducing training time by 72%.
- **Smart PDF Reader** – Built RAG system using LangChain + Llama 3.1 for large document QA.
- **Synthetic Dataset Generator** – Developed GANs for artificial synthetic dataset generation
- **Creditworthiness Prediction** – Applied ML classifiers (RF, XGBoost, SVM); achieved F1=0.88.
- **Sentiment Analysis for Amazon Reviews** – Designed LSTM-based pipeline with embeddings & TF-IDF, to classify positive and negetive reviews

Leadership Experience

- Teaching Assistant for Introduction to Computer Programming – Conducted tutorials and mentoring.
- Department Academic Mentor for 4 students, providing academic and personal guidance.
- Member, Hostel 2 Mess Committee – Managed logistics for 500+ students with INR 1.5M budget.
- Part of Placement Team – Improved industry collaborations for IIT Bombay students.
- Mentored 30+ teams across 90+ cities under E-Cell IIT Bombay.