

Akash Pothula | LLM Engineer | [Portfolio page](#)

saiakash.pothula1@gmail.com | +91-7976828261 | [linkedin](#)

EXPERIENCE

Cybernetic Controls Ltd

June 2024 – August 2025

LLM Engineer

- Automated Market Abuse Risk Assessments (MARA) using **Anthropic's Claude 3.7 sonnet model** via **AWS Bedrock** to help financial firms comply with FCA guidelines.
- Reduced manual compliance review time from 4 months to 10 minutes.
- Created a pipeline for real-time processing of uploaded compliance documents using **Amazon SQS, AWS Lambda, DynamoDB, Amazon API Gateway** and **AWS cloud formation**.
- Developed an advanced scoring algorithm using **exponential decay normalization, beta-binomial distributions** for statistical confidence, and weighted aggregation across 100+ risk factors.
- Algorithm outputs enabled firms to replace guesswork with data-driven decisions - identifying which compliance fixes prevent the most fines per pound spent reducing regulatory risk by 70%.
- Built a scalable, context-aware **custom intelligent document parser** with Claude Haiku and **AWS Step Functions**, reducing processing costs by 80%.
- Engineered a **cost-optimized RAG system** by innovatively storing embeddings in S3 hosted CSV files instead of traditional vectorDB, achieving 98% cost reduction while maintaining enterprise security through AWS.
- Implemented **KNN semantic search** from scratch using scikit-learn to retrieve relevant embeddings.
- Developed and maintained **CI/CD pipelines** using **AWS CodeBuild, ECR and Docker** to automate build, test, and deployment workflows for smoother and faster releases.

Tata Consultancy Services

Feb 2022 – Sep 2022

Assistant System Engineer

Hyderabad

- Received training in Python programming and Machine learning.

SKILLS

- **Gen AI:** LLMs, RAG, Prompt engineering, Langchain, LoRA, MLOps, OpenAI, Roberta, BERTopic
- **ML/DL:** Transformers, LSTMs, Clustering, Regression, Transfer learning, Sentiment analysis.
- **Tech Stack:** Python, SQL, Docker, Git, R
- **Libraries:** Hugging Face Transformers, boto3 SDK, Tensorflow, PyTorch, pandas, Numpy
- **Cloud/ VectorDB:** AWS, Chroma DB

EDUCATION

Indian Institute of Technology (IIT), Jodhpur

2022 - 2024

MTech in Data and Computational Sciences

- Key Courses: Natural Language Processing, Machine Learning, Software and Data Engineering, Deep Learning, Computer Vision, Artificial Intelligence, Optimization, Linear Algebra, Technology and Leadership.

PROFESSIONAL PROJECTS

- Implemented **GPT-2-small (125M parameters)** from scratch using Python and **PyTorch**. [github](#)
- Fine-tuned GPT-4o-mini on 2500 manually corrected examples for semantic topic modelling and achieved an accuracy of 95.6%.
- Built an ML-based cricket par score predictor using **web-scraped data**, where a synthetic dataset and **sigmoid function** were used to convert binary match outcomes into actionable score thresholds. [Report](#)