PRAJWAL KUMAR

Pittsburgh, PA — +1 (878) 256-0070 — prajwalk@andrew.cmu.edu — LinkedIn — GitHub — Website

SUMMARY

AI/ML Engineer specializing in Generative AI, Large Language Models (LLMs), and Scalable Data Systems. Proven track record in fine-tuning Transformers (LoRA/Peft), architecting RAG workflows with Graph Databases (Neo4j), and building production-grade ETL pipelines on AWS/GCP. Expert in Python, PyTorch, and MLOps practices (Kubernetes, Docker, Kafka) for delivering high-performance solutions in NLP and Computer Vision.

EDUCATION

Carnegie Mellon University — **Pittsburgh, PA**
Master of Science in Artificial Intelligence Engineering, GPA: 3.6 — December 2025
Coursework: Data Engineering, Deep Learning, Machine Learning in Production, Generative AI, Trustworthy AI

Maharshi Dayanand University — **Rohtak, India**
Bachelor of Technology in Computer Science and Engineering — June 2024

EXPERIENCE

Infinite Computer Solutions — **Irving, TX**
AI/ML Summer Intern — **May 2025 - August 2025**

- Engineered an Agentic AI Workflow using LangChain and Playwright to orchestrate multi-step web scraping of major telecom carriers; structured unstructured HTML data into a Knowledge Graph, reducing data aggregation time by 70%.
- Built a GraphRAG (Retrieval-Augmented Generation) pipeline using Neo4j Vector and OpenAI Embeddings, enabling enterprise teams to perform complex natural language queries over interconnected telecom data.
- Deployed the full stack (Streamlit, Neo4j, Python Agents) on Kubernetes using custom Helm Charts and Docker Compose, ensuring scalable, reproducible deployments for cross-functional analytics teams.

Qriocity — **Chennai, India**
Machine Learning Developer Intern — **January 2024 - February 2024**

- Developed a healthcare-focused ML pipeline using TensorFlow to model medical ontologies and predict prescriptions from structured clinical data, achieving 99% accuracy and demonstrating real-world applicability in clinical decision support.
- Built a Multimodal Emotion Recognition Engine using TensorFlow (DNN) and KNN; applied TF-IDF vectorization to map text to continuous Valence-Arousal-Dominance (VAD) vectors, improving empathy scoring and classification by 30%.

ScriptEdge — **Akola, India**
Macgine Learning Intern — **July 2022 - August 2023**

- Developed a text-to-image system for designer QR codes using Stable Diffusion (w/ ControlNet) from Hugging Face. Integrated patterns into user-specified images via prompts and fine-tuned diffusion parameters for optimal visual coherence.
- Engineered Bidirectional LSTM (RNN) model in TensorFlow for language identification (Text classification), 98% accuracy.

ACADEMIC RESEARCH & PROJECTS

End-to-End Movie Recommender with Kafka, Kubernetes, and A/B Testing - Carnegie Mellon (GitHub)

- Built a scalable KNN-based movie recommender (NDCG@10 = 0.9983, 65K+ QPS) with Dockerized microservices, Kubernetes, CI/CD (Jenkins), and CRON-based retraining; integrated A/B testing and real-time monitoring with Grafana.

Emotion-Aware Multimodal AI Companion - Carnegie Mellon University (GitHub)

- Built a multimodal AI companion with speech-to-text (Google API), CNN-based SER (71.2%), and Transformer models; improved LLM empathy by 30% via psychoanalysis modules and deployed a Streamlit app for real-time emotion feedback.

Malware Classification using Multi-Modal Approaches – Carnegie Mellon University (GitHub)

- Designed a multi-modal malware detection pipeline using CNN (ResNeXt), XGBoost (n-grams), and LLaMA-based LLMs, achieving 99.68% image classification accuracy and building a real-time GUI with JSON-based LLM explainability.

Cloud-Native Soccer Player Valuation Platform - Carnegie Mellon University (Github)

- Designed an end-to-end data system on GCP to predict player value; built PySpark ETL pipeline from GCS to Cloud SQL (PostgreSQL) and trained regression models (SparkML, PyTorch) using Optuna/CrossValidator, deployed via Cloud Run.

Advancing Image Security through Deep Learning and Cryptography in Healthcare Industry - IEEE (Paper)

- Conference paper on a deep learning–based cryptographic framework for securing medical images in IoMT environments.

SKILLS

Programming & Data: Python, SQL, Pandas, NumPy, PySpark, PostgreSQL, Neo4j, AWS (EC2), GCP, Streamlit
ML & GenAI: PyTorch, TensorFlow, Scikit-learn, LangChain, OpenAI GPT, LLMs, RAG, Vector Databases, Playwright
Systems & MLOps: Git, Docker, Kubernetes, Kafka, MLflow, Jenkins, CI/CD, Linux, REST APIs, Prometheus, Grafana