

PRAJWAL KUMAR

Pittsburgh, PA — +1 (412) 728-2613 — prajwalk@andrew.cmu.edu — [LinkedIn](#) — [GitHub](#) — [Website](#)

EDUCATION

Carnegie Mellon University	M.S. in Artificial Intelligence Engineering – Information Security	GPA: 3.6 / 4.0 — Dec 2025
Maharshi Dayanand University	Bachelor of Technology in Computer Science & Engineering	GPA: 8.27 / 10 — Jun 2024

MACHINE LEARNING & AI EXPERIENCE

- Engineered end-to-end Machine Learning pipelines, secure OCR-powered Flask API on Microsoft Azure Computer Vision, processing real-time data streams with Kafka, evaluating models and synthetic data generation with PyTorch, Zeno, and LLMs.
- Trained a high-F1 PyTorch intrusion detection model on GCP Vertex AI Workbench, using PySpark for data engineering. Mitigated class imbalance with a weighted loss MLP and accelerated training with GPU (MPS) and Distributed Data Parallel (DDP).
- Implemented a character-level GPT Transformer from scratch in PyTorch, architecting and comparing a novel implementation of Rotary Positional Embeddings (RoPE) against a vanilla transformer. Improved sequence generation accuracy from less than 10% to over 30% by leveraging RoPE and a full pre-training/fine-tuning pipeline.
- Designed and deployed an end-to-end data system on Google Cloud Platform (GCP) to predict player market value. Engineered a PySpark pipeline for ETL of 15 raw datasets from GCS into Cloud SQL (PostgreSQL). Trained, tuned (Optuna, CrossValidator), and compared multiple regression models (SparkML, PyTorch) before deploying the final prediction service on Cloud Run.

WORK EXPERIENCE

AI/ML Summer Intern | Infinite Computer Solutions

May 2025 – Present | Irving, Texas

- Developed an LLM-powered agentic system with LangChain, crawl4ai, and Playwright to automate customer workflows and built a Neo4j knowledge graph from interaction logs, enabling anomaly detection, reducing manual testing by 90%, and identifying friction points for out client Verizon's customer journey analysis.
- Developed a RAG system using Neo4j vector database, LangChain, and OpenAI API for semantic search and precise client query responses on customer journeys across websites; built Streamlit UI for client interaction, optimized embeddings with PyTorch, and compared Mistral LLMs, cutting query latency by 30% and improving Verizon's customer journey optimization.

Machine Learning Developer Intern | Qriocity

Jan 2024 – Feb 2024 | Chennai, India

- Engineered a deep learning-driven ontology for precise medicine prescription, leveraging TensorFlow and rdflib. Integrated structured medical data to train a neural network, achieving 99% accuracy in predicting optimal medications for diseases.
- Engineered a multi-modal mental health chatbot (text/speech) with a hybrid TensorFlow DNN and KNN model. Predicted emotional VAD scores from TF-IDF vectors, then classified into discrete emotions to provide empathetic user support.

Data Science Intern | Zummit Infolabs

Nov 2022 – Mar 2023 | Bengaluru, India

- Engineered an NLP algorithm for language identification with TensorFlow Bidirectional-LSTMs (RNN). Achieved 98% accuracy on a 1000+ sentence, 10-language dataset, demonstrating sequence modeling/text classification.
- Developed a real-time driver drowsiness detection system using Keras, achieving 94% accuracy. Trained a CNN on a 2800+ image Kaggle dataset, leveraging computer vision to classify driver states (open/closed eyes) for enhanced road safety.

ACADEMIC RESEARCH & PROJECTS

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.

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

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

End-to-End Movie Recommender with Kafka, Kubernetes, and A/B Testing - Carnegie Mellon University ([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 Prometheus + Grafana.

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

- Authored an IEEE conference paper on a novel deep learning-based cryptographic framework using chaotic systems and weight analysis to generate robust, undetectable encryption patterns for securing medical images in IoMT environments.

SKILLS

Languages & Libraries: Python, SQL, Pandas, NumPy, Scikit-learn, PyTorch, TensorFlow, Apache Spark (PySpark), Streamlit

ML & GenAI Tools: Hugging Face Transformers, LangChain, Mistral AI, OpenAI, MLflow, Vector Databases

Cloud & MLOps: AWS, GCP (Vertex AI, Gemini, BigQuery), Docker, Kubernetes, Jenkins, Prometheus, Grafana, CI/CD

Infrastructure & Databases: Git, Linux, Flask, PostgreSQL, Neo4j, Kafka, Postman