

Karthik Prakash

kprakas2@andrew.cmu.edu | LinkedIn | GitHub | +1 (412) 287-9404

EDUCATION

Carnegie Mellon University

Master of Science in Intelligent Information Systems, School of Computer Science (SCS)

Relevant Coursework: Natural Language Processing, Deep Learning, LLM Systems, AI Agents, Generative AI

Pittsburgh, PA

Dec 2027

Rashtreeya Vidyalaya College of Engineering (RVCE)

Bachelor of Engineering in Computer Science. GPA: 9.73/10.

Relevant Coursework: Data Structures & Algorithms, Database Management Systems, Linear Algebra, Probability & Statistics

Bengaluru, India

Jun 2026

EXPERIENCE

Machine Learning Intern

Draup

Jan 2026 – Jul 2026

Bengaluru, India

- Slashed Milvus vector RAM by 75% via INT8 Quantization and Matryoshka Learning, maintaining a strict 98% recall.
- Deployed a potion-base-8M on-the-fly cross-encoder, halving storage overhead and improving semantic re-ranking.
- Calibrated dynamic classifier thresholds recovering 8% false-negative signals at 83.1% precision across the retrieval pipeline.

Research Intern

EDR Evasion - Technical University of Applied Sciences (THWS), Germany

Jan 2026 – May 2026

Würzburg, Germany (Remote)

- Engineered an RL engine using Deep Q-Networks to model complex endpoint attacks as a Markov Decision Process.
- Exposed blind spots in 5 enterprise EDRs (e.g. SentinelOne), attaining a 26.6% mean detection rate and 100% evasion rate.
- Developed a scalable C++/Python adversarial platform integrating ML to automate continuous enterprise EDR red-teaming.

Software Engineering Intern

Microsoft

May 2025 – Jul 2025

Hyderabad, India

- Implemented a scalable multi-intent query processing framework for the Tenant Admin Copilot, enabling decomposition and sequential execution of complex natural language queries.
- Validated and adopted Chain-of-Thought prompting via empirical benchmarking against Multi-stage Prompt Chaining.
- Designed a metadata-driven API mapping engine using OpenAPI Swagger for seamless, real-time dependency resolution.
- Integrated a DAG-based parallel execution layer in the transpiler to orchestrate complex, multi-step concurrent API calls.

PROJECTS

ML-Driven Predictive Simulation Pipeline - Honeywell Capstone

Aug 2025 – Nov 2025

- Designed a deep learning model, accurately predicting compressor blade aerodynamic performance from 3D point cloud data.
- Reduced computational modeling costs by replacing expensive CFD simulations with fast, ML-driven surrogate predictions.
- Realized efficiency improvements of up to 9.4% through reinforcement learning-based optimization.

Intelligent Speech-Driven Information Retrieval for Agricultural Knowledge

Mar 2025 – Apr 2025

- Created a multimodal speech-to-text pipeline yielding a precise 7.08% WER for complex multilingual query retrieval.
- Achieved scalable, low-latency search recall by integrating BERT embeddings and FAISS vector search.
- Tools: Python, PyTorch, Hugging Face Transformers, Wav2Vec2-XLSR-53, Multilingual BERT, FAISS

Knowledge Graph-Enhanced RAG System for Cyber Threat Intelligence

Nov 2024 – Feb 2025

- Constructed a knowledge graph of 1,427 nodes and 2,543 relations via Neo4j and BERT embeddings to map threat data.
- Implemented a LLaMA 3.1 RAG pipeline for zero-shot content retrieval: 87.2% hit rate, 91.4% relevancy score.
- Evaluated retrieval precision via MRR and entity-recall metrics, securing a highly competitive 0.84 MRR at 4.8 entities/query.
- Tools: Python, LLaMA 3.1, Neo4j, Pinecone, BERT, PyTorch

PUBLICATIONS

[VYUHA: EDR Evasion via Deep RL-Guided Orchestration](#) - ACM CCS 2026 (Submitted, First Author)

[DL for Dementia Detection using Non-Invasive Data & Genetic Risk](#) - IEEE CMSS 2025 (First Author)

[Pulse Point: Medical Data Intelligence for Human-Centric Interaction](#) - IEEE ICOCT 2025 (First Author)

[Lightweight Machine Learning for Passive Acoustic Monitoring of Marine Mammals](#) - IEEE ICAISS 2026

SKILLS

Languages: Python, C/C++, SQL, R, Bash

Machine Learning: PyTorch, TensorFlow, Scikit-Learn, OpenCV, XGBoost

Data Science & Visualization: Pandas, NumPy, SciPy, Matplotlib, Seaborn, Tableau

GenAI & NLP: Large Language Models (LLMs), RAG, LangChain, Hugging Face, Vector Databases

MLOps & Cloud: AWS, GCP, Docker, Kubernetes, Linux, Git, MLflow, CI/CD