Vikraman Parthiban

Email: pvikraman86@gmail.com | Linkedin: linkedin.com/in/vikraman-p-332549176/

GitHub: github.com/infinity-void6

Education

SRMIST Ramapuram,Chennai 2021 - 2025
BTech in Computer Science Engineering with AI&ML specialization

Technical Skills

Languages: Python, JavaScript, C++, SQL

DS/ML: PyTorch, Numpy, sklearn, pandas

Tools: Git, GitHub, VSCode,Linux

Robotics: ROS2, Isaac Sim

Databases: SQL (Postgre SQL),Neo4J, SQLAlchemy, MilvusDB

Professional Experience

IIT Madras (Indian Institute of Technology Madras) May 2025 - Ongoing
Project Intern Chennai, India (On-site)

- Built an end-to-end Retrieval-Augmented Generation (RAG) based course assistant chatbot for NPTEL courses, evolving an early proof-of-concept into a production-oriented system supporting large-scale, self-paced learning.
- Designed a scalable retrieval pipeline by consolidating textbooks and YouTube lecture transcripts using a novel cross-source chunking strategy to improve semantic relevance and answer quality.
- Migrated the system from a prototype stack (Dash, Ollama, ChromaDB) to a production backend using FastAPI, vLLM, and Milvus, enabling multi-user concurrency and reliable deployment.
- Optimized inference and retrieval performance, reducing average response latency from $\sim$80 s to $\sim$13 s under 100 concurrent requests while supporting 20+ simultaneous users.

IISER Bhopal (Indian Institutes of Science Education and Research) May 2023 – July 2023
Research Intern- Data Science Department Bhopal, India (On-site)

- Built knowledge graphs to explore potential drug–disease associations for repurposing from biomedical literature using dataset from Zitnik Lab's PrimeKG dataset.
- Conducted literature review on AI-driven drug discovery, focusing on knowledge graph applications and NLP methods to identify key industry leaders and understand the methods they utilized.

Projects

Multimodal-Course Assistant Chatbot for Control Theory, Virtual Assistant and Law August 2025

- Developed advanced chatbot features including automated question answering, follow-up question generation, mind-map style visualizations, and PDF extraction and summarization for course content.
- Implemented and compared multiple LLM backends (LLaMA, Qwen, Gemini) to study trade-offs between response quality, inference latency, and token efficiency in a real-world educational setting.
- Experimented with structured output formats (JSON vs. YAML) to reduce token overhead while maintaining deterministic and machine-readable responses.
- Conducted systematic concurrency and stress testing using `k6` and Postman to identify inference and retrieval bottlenecks and guide performance optimizations.
- **Tools Used**: Python, LangChain, FastAPI, vLLM, Ollama, MilvusDB, ChromaDB, Linux, Bash, k6, Postman

Anomaly Detection in Surveillance Videos using Multiple Instance Learning and Weakly Supervised Learning December 2024

- Designed an AI-powered anomaly detection system under the guidance of Dr. Prasanna Balaprakash (Director of AI Programs at Oak Ridge National Laboratory) for real-time surveillance, reducing the need for manual

monitoring.

- Implemented real-time video feature extraction using Meta's pretrained 3D CNN, designed a Deep GRU-based network, and optimized training efficiency with mixed precision on local GPUs.
- comparative study between I3D and X3D results on surveillance data
- **Tools Used**:PyTorch, Streamlit, Numpy

Self-Supervised Vision Transformer for Pathology August 2024

- Implemented a self-supervised learning framework under Dr. Krishna Chaitanya (Research Scientist at Johnson and Johnson) to enable efficient pathology image analysis for disease detection, reducing dependency on labeled medical data.
- Used Vision Transformers (DINO) for feature extraction, enabling classification and clustering of pathology images.
- Improved feature representation, leading to better downstream classification performance in pathology datasets.
- **Tools Used**: PyTorch, Matplotlib, scikitlearn

Knowledge Graph for Rare Diseases May 2023

- Developed knowledge graphs for rare diseases (e.g., muscular dystrophy, Alzheimer's) under Professor Parthiban Srinivasan (Director of AI at AVMC) to enhance drug discovery and biomedical research
- Extracted and processed data from scientific literature using ETL pipelines, formatted it for graph-based representation, and integrated it with Neo4J for efficient querying.
- Improved disease-drug relationship retrieval, enabling faster identification of potential drug repurposing candidates.
- **Tools Used**: Neo4J, pandas, Python

Conferences and Workshops

ROSCon 2025, Singapore October 2025

- Attended ROSCon 2025 with hands-on exposure to cutting-edge robotics software and tooling.
- Completed the Isaac Sim Workshop focused on simulation-driven robotics development and testing.
- Participated in the "How to Implement a Full ROS 2 Application: a Tic-Tac-Toe Player Robot" workshop, integrating perception, motion planning, and an AI engine in ROS 2.
- Learnt the basics of ROS2, including nodes, topics, publishers–subscribers, and multi-package workspace setup.

PyCon India, Bengaluru September 2024

- Presented a poster on "Generative AI for Drug Discovery."

ETH Zürich Machine Learning Summer School for Healthcare August 2024

- Engaged in AI applications in healthcare with industry, government, and academia experts
- participated in lectures, tutorials, and discussions that involved tackling bias in datasets affecting AI models.
- Explored foundation models in healthcare and discussed paradigm shifts for technology adoption in healthcare.