NAVEENRAJ KAMALAKANNAN

naveenraj.k@nyu.edu | Brooklyn, New York | +1 914-490-3063| Personal Website | GitHub | LinkedIn

PROFESSIONAL EXPERIENCE

J.P. MORGAN — **New York, United States**
AI and Data Science Associate - Intern — **June 2025 - August 2025**

- Architected and enhanced Asset and Wealth Management's agentic platform, improving the relevance and naturalness of AI-generated content through a multi-stage retrieval pipeline.
- Orchestrated multiple AI agents with advanced reasoning capabilities to significantly enhance existing product functionality.
- Applied expertise in OpenSearch, RAG Systems, LangChain, KV Cache Optimization, and MCP Server to develop robust and performant AI solutions for financial applications.

NYU CENTER FOR DATA SCIENCE & NYU LANGONE — **New York, United States**
Research Assistant — **February 2025 - May 2025**

- NSF-funded research role with NYU Center for Data Science and NYU School of Medicine focused on advancing stroke rehabilitation.
- Investigated and benchmarked state-of-the-art video understanding models (InternVL2, NVLA, LLaVa OneVision) for precise sub-second action detection in stroke rehabilitation scenarios.
- Engineered a multi-stage pipeline integrating YOLO11 for human-pose detection, DINOv2 Prompt based Object Detection and OpenMMLab's pose estimation algorithms for motion analysis.
- Developed and integrated a Visual Language Model (VLM) based on video frames for improved contextual detection of patient interactions with objects.

ZEECO MIDDLE EAST — **Dammam, Saudi Arabia**
Control Engineer — **July 2022 - August 2023**

- Engineered a custom anomaly detection system for circuit designs and equipment datasheets using neural networks, resulting in a 40% reduction of man-hours.
- Formulated a predictive model to optimize Pressure, Temperature, and Flow controllers, achieving 30% increased efficiency in process control, leading to significant cost savings and reduced man-hours.

BAJAJ FINSERV — **Pune, India**
Data Engineer - Intern — **January 2022 - June 2022**

- Streamlined Azure-based data migration pipelines, reducing migration time by 34% and saving operational costs during the transition from EDW to Cosmos DB, leveraging Data Factory and Data Lake.
- Developed an ML model, leading to a ~33.3% reduced Data Warehouse Units resource consumption in Azure EDW, optimizing resource utilization and identifying non-optimal queries, leading to significant cost savings.

EDUCATION

NEW YORK UNIVERSITY — **New York, United States**
MS in Computer Engineering - GPA: 4.00 / 4.00 — **September 2024 - May 2026**

- Explored Foundational Vision Language Models and Human Pose Estimators
- Collaborated in a team to build Multi-Modal Retrieval Systems & fine-tuned models for Video Understanding
- Relevant Coursework: Deep Learning, Reinforcement Learning, Optimization.

THE CONSTRUCT — **Barcelona, Spain (Remote)**
Robotics Developer Bootcamp - Score: 97.4% — **September 2023 - March 2024**

- Mastered implementing algorithms such as A* in Nav2 stack, RRT in OMPL, IK-FAST in trajectory optimization, and gained hands-on experience in SLAM, and vSLAM. Ranked within the TOP 1% of the batch.

VELLORE INSTITUTE OF TECHNOLOGY — **Vellore, India**
Bachelor of Technology in Electronics and Communication - GPA: 8.79 / 10.0 — **July 2018 - May 2022**

- Relevant Coursework: Deep Learning, Machine Learning, Operating Systems, Linear Systems.

OPEN SOURCE CONTRIBUTIONS

- **DeepSpeed by Microsoft:** Resolved a gradient norm bug in the ZeRO-3 optimizer for CPU offload and ensured its stability by implementing comprehensive unit tests for FP16/BF16. - **[GitHub PR]**
- **ArcticInference by Snowflake:** Engineered FlashInfer backend support into SwiftKV, boosting throughput by ~49% over FlashAttention while debugging metadata handling to maintain model accuracy - **[GitHub PR]**

PROJECTS

AUTOMATED CHEST X-RAY PERCEPTION MODEL **November 2021 - January 2024**

- Collaborated with a team of three scholars to develop a diagnostic model introducing Exponential Pixelating Integral transforms, fractals, and an adaptive splines-based model, resulting in an overall accuracy of 99.63%.
- Achieved accuracy and F1 metrics, ranging from 98.46% to 99.45% and 96.53% to 98.10% respectively, utilizing the Multivariate Adaptive Regression Splines (MARS) ensemble model.

ANALYZING BRAIN WAVES FOR DIAGNOSIS: EEG SIGNALS **January 2021 - May 2021**

- Led a team to develop a Deep Learning model analyzing EEG signals, achieving 93.3% accuracy in detecting early-stage Parkinson's disease.
- Attained an F1 score of 93.48% and presented findings at the University of Tubingen Symposium.

EARLY DETECTION OF SEPSIS - NATIONAL HACKATHON **December 2019 - March 2020**

- Led a cross-functional team to develop a sepsis onset detection model using critical biomarkers (PCT and MDW), securing first place in the Design Category and winning a $2,000 grant from the VIT Incubator

PUBLICATIONS

- **Kamalakannan N**, Macharla S, Kanimozhi M and Sudhakar M S. "Exponential Pixelating Integral Transform with Dual Fractal Features for Enhanced Chest X-Ray Abnormality Detection" Computers in Biology and Medicine, Volume 182, 2024 - **[DOI]**
- **Kamalakannan, Naveenraj**, Shiva Prasaath Sudha Balamurugan, and Kalaivani Shanmugam. "A novel approach for the early detection of Parkinson's disease using EEG signal." IJEET 12.5 (2021): 80-95 - **[DOI]**

SKILLS

- **AI/ML:** Machine Learning, Deep Learning, Reinforcement Learning, LLM Fine-Tuning, Vision-Language Models, Mixture-of-Experts, PyTorch
- **Accelerated Computing:** CUDA, Triton, OpenMP, cuDNN, CUTLASS, GPU Kernel Optimization, KV Cache Optimization
- **Tools:** MLFlow, Tensorboard, Ray, ONNX
- **Programming Languages:** Python, Go Lang, C/C++
- **Cloud & Infrastructure:** AWS, Azure, GCP, Kubernetes, Docker, Jenkins, Spinnaker
- **Databases & APIs:** SQL, FastAPI, Firebase