AARNAV HARIRAMANI
(732) 604-2533 | aarnavhariramani@gmail.com | Monroe Township, NJ
LinkedIn | GitHub

EDUCATION

Stevens Institute of Technology **Hoboken, NJ**
B.S. in Computer Science, **GPA:** 4.0/4.0 *August 2024 – May 2027*
Relevant Coursework: Algorithms, Data Structures, Discrete Structures, Linear Algebra, Vectors and Matrices, Multivariable Calculus, Probability and Statistics, Intermediate Statistics, Computer Architecture and Organization, Mechanics
Honors and Activities: Delta Sigma Pi Business Fraternity (Brother), Dean's List (1st Semester, 2nd Semester), Stevens Blueprint (Technical Developer), Stevens Computer Science Club (Member), Stevens SASA (South Indian Student Association)

EXPERIENCE

eBay x Stevens Institute of Technology **Hoboken, NJ**
Undergraduate Researcher *December 2024 – Present*
- Conducting research with Professor Jia Xu in collaboration with eBay Research, optimizing Large Language Models (Llama 3, GPT-4, and Mistral) for more accurate evaluation of LLM produced data (LLM-as-Judge), focusing on prompt and train set refinement.
- Proposed and programmed NEATJudge, a first-of-its-kind evolutionary algorithm using NeuroEvolution of Augmented Topologies (NEAT) to evolve a feed forward neural network data and prompt selector, improving the judge model's accuracy by ~23%.
- Authored and presented a research paper: "What Test is the Right Test?" at eBay Research's internal research conference, while preparing for external submission with the International Conference on Learning Representations (ICLR, <5% acceptance rate).

Stevens Institute of Technology - Student Managed Investment Fund (SSMIF) **Hoboken, NJ**
Quantitative Analyst Intern *January 2025 – Present*
- Developed a multi-factor model within the Macro Strategies team (specifics under an NDA), to identify macroeconomic factors with strong predictive power that forecast financial sector returns for an investment fund managing over $1.5M in total assets.
- Enhancing an alpha generating model's predictive capabilities by leveraging seasonal decomposition and Fast Fourier Transform (FFT) to diagnose residuals for independence and normality (normally distributed) for robust sector return forecasting accuracy.
- Co-designed and implemented new predictive and machine learning models (OLS, CAPM, HAR, RNN) and built a centralized model registry system to streamline experimentation, version control, and deployment processes across the Macro Strategies pipeline.

Aquatics United **Princeton, NJ**
Website Development Intern *December 2023 – February 2024*
- Redesigned, optimized, and modernized a website's front-end using HTML, CSS, and React, improving mobile navigation, which boosted mobile traffic by 15%, and integrated Fetch API with backend systems to reduce response times to ~200 ms on average.
- Enhanced search ranking and indexing by building an SEO automation pipeline with BeautifulSoup and spaCy for meta tag generation and refining crawler access via robots.txt and XML sitemaps leading to a 17% increase in indexing/ranking on Google.
- Implemented a machine learning model with scikit-learn to analyze user traffic logs and predict peak usage patterns, enabling proactive caching that reduced average page load time by 120 ms and improved scalability under sustained high traffic.

PROJECTS/AWARDS

FAISS-Powered Semantic RAG Chatbot *June 2025 - Present*
- Engineered a FAISS (Facebook AI Similarity Search)-based semantic search system indexing 49,470 eBay product records with all-MiniLM-L6-v2 (384-dimension) embeddings, compressing the index to ~72.5 MB and enabling sub-100 ms vector retrieval.
- Integrated retrieval with a conversational frontend by deploying a Streamlit RAG chatbot integrating the FAISS search system for semantic retrieval and Llama-3.3-70B-Instruct-Turbo for conversational product search, returning relevant purchase information.
- Optimized memory usage and retrieval efficiency by applying product quantization for index compression and parallelized query batching, reducing overall computational overhead by 34% and ensuring cost effective scalability in deployment environments.

Cognate Detection Using Siamese Networks *December 2024 - Present*
- Writing a research paper exploring how integrating the Siamese architecture into cognate language detection, captures the linguistic similarities of cognates via cross embeddings. The goal is to train models for phonetic similarities common to cognates.
- Fine-tuned Qwen 2.5 using PyTorch & HuggingFace Transformers for NLP tasks, incorporating cultural sensitivities from region-specific corpora; improved BLEU score by 12% via transfer learning and boosted engagement metrics by 15% in pilot evaluations.

SKILLS AND INTERESTS

Programming Languages: Python, R, Java, JavaScript, C/C++, SQL, HTML/CSS, PHP, Swift
Machine Learning and Data Science: PyTorch, TensorFlow, NumPy, Keras, Ollama, HuggingFace, Pandas, scikit-learn, FAISS, spaCy, Selenium, Natural Language Processing
Tools and Systems: AWS, Docker, Firebase, Git, MySQL, SQLAlchemy, Stripe API, REST/Fetch API, Terminal