Inesh Chakrabarti

858-925-3059 | inesh33@g.ucla.edu | linkedin.com/in/inesh-chakrabarti | github.com/beesfleas

Education

University of California, Los Angeles Los Angeles, CA

B.S. Electrical Engineering, M.S. Electrical Engineering *Expected Graduation – June 2027*

- **Graduate GPA:** 4.0 / 4.0, **Undergraduate Major GPA:** 3.8 / 4.0
- **Coursework:** Large Scale Data Mining, Convex Optimization, Deep Learning, Software Engineering, Embedded Systems, Computer Architecture, GPU Microarchitectures, Numerical Computing, Stochastic Systems, Communications, Signals and Systems, Probability and Statistics
- **Societies:** American Nuclear Society (President, Founder), Eta Kappa Nu (Mentorship Chair)

Experience

UCLA Lin Yang Research Group *Febuary 2025 - Present*

- *NoWag: A Unified Framework for Shape Preserving Compression of Large Language Models* Lawrence Liu, **Inesh Chakrabarti**, Yixiao Li, Mengdi Wang, Tuo Zhao, Lin F. Yang Publication accepted to **COLM** and **ICLR SLLM Workshop**
- Built dequantization/inference kernels in C (CUDA) for parallelization over multiple GPUs while using **48x less calibration data** and maintaining performance against SOTA VQ methods
- Implemented Trellis Quantization and benchmarking in Python for NoWag, a set of shape-preserving pruning and quantization algorithms for LLMs

UCLA Complex Networks Group (Paid Student Researcher) *February 2022 - June 2024*

- Implemented High Frequency Oscillation Detector using Variational Autoencoder for neural signals, doubling number of detections with only a 10% increase in false positive
- Constructed a speech to test pipeline that subtitled recall experiements with precise temporal acc.
- Processed and visualized neural spike data using Python and MATLAB to demonstrate correlation between individual neural spikes and character recognition from animation
- Developed a **pipeline** for EEG data analysis with wavelet transform pre-processing and model distillation to predict human movement using transformer, LSTM, and CNN models.

Projects

Reinforcement Learning Hearts *September 2024 - January 2025*

- Created RL agent for Hearts using **Counterfactual Regret Minimization** and **Monte Carlo Tree Search** that reaches approximate Nash Equilibrium.
- Enhanced the Hearts project with a Tkinter UI and collaborated in a 3-person team, providing a **real-time interface** allowing for physical gameplay simulation via computer vision.

Algorithmic Trading Bot *March 2023 - May 2023*

- Developed and deployed ARIMAX-based trading bot utilizing time-series analysis to optimize trading strategies and participate in **live trading** against other bots and consumers.
- Created portfolio optimization bot with Kernel Density Estimation, achieving a **1.4 Sharpe Ratio** and **highest profit** in competition.

Large Scale Data Mining *January 2023 - March 2023*

- Experimented with different forms of clustering for text and image data including Kmeans and HDBSCAN clustering Recommender Systems
- Constructed an end to end pipeline for news classification using gridsearching over different vecotrization and classification models including BOW, TFIDF, SVM, Naive Bayes, etc. while leveraging cuML and cuDF for GPU optimization.

Skills

- **Programming Languages:** C, C++, Python (NUMBA, PySpark, Matplotlib, PyTorch, Pandas, Keras, Tensorflow), Triton, SQL, x64, C#, Java, MATLAB, R, JavaScript
- **Tools:** Docker, Git, LangGraph, MongoDB, LTSpice, GDB, Unix Shell, CUDA, OpenMP, Joblib, Django, NVIDIA Nsight Compute, Apache Spark, Fuzzing (AFL), CI/CD