Shlok Channawar

(814)-769-6825 shlokchannawar05@gmail.com linkedin.com/in/shlok-channawar/ github.com/shlok1808

Education

The Pennsylvania State University State College, PA
Bachelor of Science in Applied Data Science *December 2027*

Technical Skills

Languages: Python (Pandas, NumPy, Matplotlib, Seaborn), R, SQL
Tools & Frameworks: Git, Jupyter Notebook, Scikit-learn, Pytorch, Tensorflow, APIs, Hugging Face, VS Code

Projects

Research on Mechanistic Interpretability in LLMs | *Python, PyTorch* Spring 2026

- Proposed and validated a framework showing SAE decoder-space geometry (neighbor density, max cosine similarity) predicts feature steerability before any intervention, achieving Spearman ρ up to -0.546 ($p < 10^{-6}$) and AUROC up to 0.822
- Replicated geometry–steerability signal across 4 model architectures (Gemma-2 2B/9B, Llama-3.1-8B, Qwen3-8B) and 2 SAE widths (16K/65K); identified depth-boundary conditions where signal degrades
- Designed end-to-end evaluation pipeline covering 100+ features $\times$ 100 prompts $\times$ 9 steering coefficients with coherence filtering, SALADBench refusal scoring, and permutation null tests
- Optimized batched inference to reduce model calls by 94% (69K $\rightarrow$ 4.3K), cutting GPU compute cost on Lambda Cloud
- Led 4-person research team across pipeline engineering, ablations, and experiment tracking; submitted to a top ML workshop

SafeQuant: Quantization Safety Diagnostics for LLMs | *Python, PyTorch* In Progress (Spring 2026)

- Identified "read-side collapse" as the dominant failure mechanism in quantized LLMs: in 10/16 conditions across Llama-3.1-8B, Mistral-7B, and Qwen2.5-7B, restoring attention input projections recovered more safety than output projections
- Demonstrated full reversibility of quantization safety damage via layer-swap ablation across 3 models $\times$ 6 quantizers (INT8, NF4, GPTQ, AQLM-2bit), recovering ASR to within 4pp of FP32 baseline in all 16 conditions
- Developed a training-free delta correction fix ($W_{safe} = 0.25 \cdot W_{2bit} + 0.75 \cdot W_{INT8}$) recovering 90–95% of lost safety on Llama-3.1-8B/AQLM-2bit with no retraining
- Evaluated on WildJailbreak (100 prompts, HarmBench classifier) and ARC-Challenge; targeting EMNLP 2026

Multi-Asset Financial Portfolio Risk Analysis | *Python, Finance* June 2025

- Built a three-stock portfolio (Morgan Stanley, JPMorgan, Bank of America) and compared daily/simple/log/cumulative returns to analyze behavior and growth over time
- Assessed risk/performance using Alpha, Beta, Sharpe, Sortino, Treynor, stdev, max drawdown, VaR, and CVaR to quantify risk-adjusted returns and tail risk

Experience

Research Assistant January 2025 – June 2025
Humanitarian Engineering and Social Entrepreneurship – Hexarech *State College, PA & Kisumu, Kenya*

- Identified the need for sensor data integration in cage designs to improve farming practices, focusing on factors like fish mortality and inefficiencies
- Analyzed sensor data to identify trends, enabling farmers to adjust feed schedules and mitigate unfavorable water conditions, reducing fish mortality rates by 30–50%
- Traveled to Kenya to assess on-ground challenges, leading to piloting practical, low-cost demonstration cages tailored for smallholder farmers
- Collaborated with local stakeholders—including BMUs, hatchery owners, and extension officers—to co-design scalable cage systems based on locally available materials and investment models