

KALASH KANKARIA

+91 9314027008

github.com/Kalash0309

linkedin.com/in/kalash-kankaria/

kkankari@andrew.cmu.edu

EDUCATION

CARNEGIE MELLON UNIVERSITY, TEPPER SCHOOL OF BUSINESS Pittsburgh, US
Master of Science in Computational Finance – MSCF 12/27

- **Upcoming Coursework:** Stochastic Calculus, Options, Investments, Fixed Income, Simulations for Option Pricing.
- **Achievements:** Squarepoint Fellow 2026-27 (USD 20,000 scholarship).

INDIAN INSTITUTE OF TECHNOLOGY Gandhinagar, IN
B.Tech in Computer Science and Engineering 06/24

- **Coursework:** Probability, Statistics, Linear Algebra, Calculus, Algorithms, Machine Learning, Probabilistic ML, NLP.
- **Achievements:** Dean's List in four semesters for academic excellence, NTSE Scholarship (INR 1,20,000) by the Central Govt. of India, Excellence Scholarship in Sports and Games.

SKILLS

- **Programming:** Python, C++, SQL, Linux/Unix, Bash, NumPy, Pandas, Pytorch, JAX, Scikit-Learn, SciPy, Docker.
- **Knowledge:** Machine Learning, Probabilistic Modeling, Statistical Learning, Large-Scale Data, Time-Series, Hypothesis Testing, Model Validation, Signal Extraction, Feature Engineering, NLP, Simulation-based Evaluation.
- **Certifications:** Akuna Capital Options 101, Deep Learning Specialization by Andrew Ng.

EXPERIENCE

SKAN AI (AI-powered process intelligence platform for enterprise operations) Bangalore, IN
Data Scientist II 03/26 - 06/26

- **Computational Optimization:** Redesigned LLM driven feature extraction engine with Vector DB-backed RAG pipeline, achieving LLM calls reduction by 35%; quantified through hyperparameter-tuned simulation framework.
- **Model Orchestration:** Integrated multi-LLM inference layer leveraging LiteLLM, enabling unified access to local Ollama models and external APIs via load balanced request routing and asynchronous distributed inference.

Data Scientist I 07/24 - 02/26

- **Structural Screen Clustering:** Modeled latent webpage structure from HTML DOM, separating stable layout signals from instance-level noise to cluster and classify screens with 87% accuracy, enabling diverse subset sampling.
- **Hierarchical Activity Discovery:** Modeled noisy sequential UI events into hierarchical activity representations, abstracting granular actions into composite multi-step workflows with linguistic and metadata enrichment.
- **Text Mining:** Pioneered a low-latency agentic workflow for email feature extraction and process classification, attaining 94%+ accuracy and reducing observation gaps by up to 50% for investment banking clients.
- **Intent Modeling:** Built unsupervised semantic clustering pipeline for email action summaries, normalizing them into canonical intents and broader categories with 68% accuracy, supporting dynamic classification of new intents.
- **Model Benchmarking:** Researched transformer architecture models for event grouping and task identification utilizing fine-tuning and transfer-learning on large datasets, averaging 94% similarity score.

PROJECTS

Bayesian Unlearning - Learning to Forget | [\[Links\]](#) 08/23 - 04/24

- **Probabilistic Unlearning:** Investigated methods for selective data removal without retraining, using closed-form updates and Laplace, Variational Inference, and Markov Chain Monte Carlo approximations for posteriors.
- **Hypothesis Testing:** Achieved exact posterior recovery for linear regression on deleting up to 50% of training data.
- **Computational Analysis:** Assessed Laplace approximation on nonlinear classification tasks, accomplishing 75–79% retraining-gap closure and 2.0–2.9× faster updates than retraining under 5% targeted deletion.
- **Image Unlearning:** Extended unlearning concepts to vision models employing methods such as fine-tuning resets, noise-injection fast unlearning, and layer re-initialization, on CelebA_HQ dataset; Analysed with GradCAM.

Neural Networks vs Tree Based Algorithm on Tabular Data | [\[Report\]](#) 09/23 - 11/23

- **Comparative Modeling:** Benchmarked neural network architectures against XGBoost and other tree-based models on tabular datasets, identifying regimes where each model class generalized better.
- **Representation Learning:** Applied feature rotation, smoothing, embeddings, augmentation, and regularization to improve deep learning performance; evaluated Hopular and TabPFN against tree-based baselines.

ADDITIONAL INFORMATION

- **Awards:** Gold Medal for Overall Outstanding Performance in Sports; Gold Medal for Outstanding Performance in Aquatics; Bronze Medal in Inter IIT Aquatics Meet 2023.
- **Leadership:** Captain, Aquatics Varsity; Vice Captain, Basketball Varsity; Teaching Assistant, Machine Learning ES335.
- **Interests:** Poker, Chess, Matiks.