Sahibnoor Singh Chahal

More About Me | csahibnoor@gmail.com | linkedin.com/in/sahibnoorsingh | github.com/Sahibnoor

Research Interests

ML Systems and GPU kernel optimization, with emphasis on using fine-grained hardware profiling signals like SM occupancy, memory bandwidth, and register pressure to guide LLM-driven automated kernel generation instead of black-box timing feedback.

Education

University of Massachusetts, Isenberg School of Management | *Amherst, MA* Sep. 2023 — Dec. 2024
Master of Science in Business Analytics *(GPA: 3.78)*

Bennett University | *India* Aug. 2016 — Jul. 2020
Bachelor of Technology in Computer Science *(GPA: 3.36, Specialization in AIML)*

Selected Research Projects

Protein Sequence Classification Using CNN + Evolutionary Optimization

- Proposed Genetic-ProtCNN, which evolves ProtCNN-style convolutional architectures with a genetic algorithm. Candidate CNNs are assembled from predefined building blocks, trained, ranked by accuracy, and recombined over 50 iterations.
- Framed protein family annotation as multiclass sequence classification on Pfam v32.0 (17,929 families), using one-hot encoded, padded amino acid sequences.
- Validated on two splits: a random 80/10/10 split and a clustered split built for distant homologs. Reported error by each test sequence's identity to the training set, to avoid inflated results from evolutionarily related sequences.
- Benchmarked against BLASTp, profile HMM, ProtCNN, and the ProtENN ensemble (13 ProtCNNs). Overall error: Genetic-ProtCNN 11.5%, ProtENN 12.2%, HMM 18.1%, ProtCNN 27.7%, BLASTp 35.9%.
- Outperformed alignment-based methods in the 30 to 90% identity range. It led in several low-identity bins on the random split, and in the 13 to 21% bins on the clustered split while staying close to the other deep models.

Optimization of Neural Network Architecture using Genetic Programming for classification of Breast Cancer

- Developed GONN, a genetic programming approach that evolves neural network classifiers. Function set: P, W, +, -, *, %. Terminals: the dataset's feature variables plus a random constant.
- Designed a hybrid crossover that sends part of each generation through hill climbing for faster convergence and the rest through standard crossover to reduce local maxima. Each generation is split into reproduction (top Pr%), mutation (bottom Pm%), and crossover (remaining Pc%).
- Implemented in Python and evaluated on the Wisconsin Breast Cancer Database (699 samples, benign vs malignant) with 50-50, 60-40, and 70-30 train/test splits.
- Reached best accuracies of 97% (50-50) and up to 99% (70-30), with confusion matrices reported per split. Results are competitive with published methods cited in the paper (C4.5 94.94%, fuzzy-genetic 97.36%, genetic programming 89.63%).

Categorizing Different Types of Online Harassment Language on Social Media

- Designed a neural text-classification system for identifying multiple harassment categories (binary + 3 subtype labels) using embeddings, CNN layers, and recurrent components.
- Processed and cleaned nearly 25k social media text samples with tokenization, normalization, and multi-label encoding.
- Implemented threshold tuning and class-weighting strategies to improve performance under label imbalance.
- **Results:** Improved multi-label F1 scores by approximately 8–12% over baseline statistical models and achieved stable performance on noisy, real-world Twitter data.

APRIL - A Program Reciting Intelligent Lyrics

- Built LSTM-based generative models trained on 5,000+ lyric sequences and 500+ MIDI files, incorporating dataset-specific pre-processing for text and audio.
- Implemented temperature-controlled sampling, sequence-windowing, and vocabulary management to reduce model collapse.
- Combined lyric-generation and melody-generation architectures into a unified generative pipeline.
- **Results:** Produced stylistically coherent lyric and melody sequences with noticeable structural similarity to the training corpus and demonstrated stable generation across multiple sampling temperatures.

Emotion Recognition from EEG Signals (FFT + Genetic Programming)

- Built a full EEG-processing workflow using Fourier-based frequency decomposition on 20-participant recorded EEG datasets.
- Engineered Genetic Programming classifiers and tuned fitness functions for multi-class emotional states (4 classes).
- Benchmarked GP-based models against SVM, Random Forest, and Neural Network baselines.
- **Results:** Achieved 89% average classification accuracy, outperforming classical ML baselines by 6–10 percentage points across emotional-state categories.

Research Experience

Independent Researcher | *GPU Kernel Optimization Project* Apr. 2026 – present

- Investigating whether hardware profiling signals (Nsight Compute: occupancy, memory throughput, warp stall reasons) can guide kernel tuning in Triton, using a shared harness with correctness checks against PyTorch eager and torch.compile on an RTX 3060 (6 GB).
- Built Triton kernels (softmax, tiled matmul, fused bias+GELU+dropout). Profiling showed softmax memory-bound (91% memory throughput) with occupancy capped at 33% by register pressure. A tiled online-softmax rewrite raised occupancy to 97% but ran 73% slower from an extra memory pass, documented as a negative result.
- Found via ncu that fp32 matmul ran without tensor cores. TF32 engaged them for a 45% speedup over eager. Fused bias+GELU+dropout reached 3.3x, matching the predicted 6-to-2 memory pass reduction.
- Classified each op as memory- or compute-bound with achieved vs. peak throughput. Softmax reached 1.9x over eager at 8192×32768.

Research Intern | *Dept. of Computer Science Engineering, Bennett University* Jan. 2020 – Jun. 2020

- Contributed to ongoing lab research on EEG-based emotion recognition (FFT feature extraction, genetic-programming classification) in Dr. Garg's group.

Summer Research Immersion | *Georgia Institute of Technology, USA* Jul. 2017

- Developed image-processing and enhancement filters in Python; worked on embedded computing and object-detection tasks using the NAO humanoid robot platform.

Professional Experience

Data Engineer (Contract) | *PRN Pharmaceuticals* Dec. 2025 – Feb. 2026

- Reconciled duplicate patient and practice records across **9 disparate acquired-practice data sources** into a consolidated NetSuite database following company acquisitions, resolving inconsistencies from manually entered sales/marketing data.
- Built a blocking-indexed fuzzy matching pipeline in Python (`rapidfuzz`) with field-level normalization (name, address, phone, email) to score and classify candidate matches at scale without brute-force pairwise comparison.
- Built a synthetic data generator with controlled duplicate injection and realistic corruption (typos, formatting noise, case variation) to create labeled ground truth, then trained a Random Forest classifier on similarity features to validate and tune matching accuracy.
- Delivered a resolved, deduplicated dataset with source IDs mapped back to consolidated partner records for downstream use.

Team Lead - Outreach & Fundraising | *Bright Mind Enrichment & Schooling* Feb. 2025 - Feb 2026

- Led a cross-functional volunteer team of 40+ members; designed a donor-tracking system using engagement scoring to prioritize outreach.
- Oversaw a parallel team coordinating with standardized data collection, data storage and reporting workflows to support outcome-driven execution across a multi-channel fundraising operation.

Data Engineer | *GagnaPro Solutions* Dec. 2020 – Jun. 2023

- Drove **$7.5M+ in incremental revenue** by engineering churn-segmentation pipelines on **14M+ user records**, enabling hyper-targeted retention campaigns.
- **Reduced data refresh latency by 40%** and improved cross-team availability by building 20+ automated pipelines with Python, Airflow, and AWS (EC2/Redshift).
- Cut report generation time by 37% through designing scalable Snowflake data marts and optimizing complex SQL workflows.
- Accelerated experimentation cycles by 30% through **PySpark optimizations**, enhancing model scalability.

Skills

GPU/Parallel Programming: Triton, CUDA, Nsight Compute/Systems, PyTorch (`torch.compile`), kernel profiling, roofline analysis

Machine Learning: Regression, Classification, Gradient Boosting, Random Forest, Time Series Forecasting, Model Optimization

Programming: Python (pandas, numpy, sklearn, PySpark), SQL (ETL, CTE, Window Functions), PSQL, DAX (PowerBI)

Data Systems: Databricks, AWS Redshift, Snowflake, Airflow

Certifications: *MITx 6.86x: ML with Python* | *IBM Data Analyst* | *IBM Machine Learning Professional*

Education

University of Massachusetts, Isenberg School of Management | *Amherst, MA* Sep. 2023 — Dec. 2024
Master of Science in Business Analytics *(GPA: 3.78)*

Bennett University | *India* Aug. 2016 — Jul. 2020
Bachelor of Technology in Computer Science *(GPA: 3.36, Specialization in AIML)*