

Sajal Gupta

✉ sajal.gupta@colorado.edu  [Website](#)  github.com/sajalgupta0422  linkedin.com/in/sajal-gupta

Education

University of Colorado Boulder

Aug 2020 – Present

PhD, Astrophysics (GPA: 3.92 / 4.00)

Boulder, Colorado

- **Publications:** Two first-author peer-reviewed papers on MPI-parallelized, global-scale relativistic plasma modeling, optimized for high-performance computing clusters (arXiv:[2409.09165](#), [2601.16370](#))

Indian Institute of Science Education and Research Kolkata

Aug 2015 – May 2020

Bachelor and Master of Science, Physics (GPA: 3.7 / 4.00)

Kolkata, India

- **Award:** Merit-based scholarship (INSPIRE)
- **Publications:** 4+ first-author publications on highly efficient simulation of astrophysical objects ([Scholar list](#))

Experience

Graduate Research Assistant

Aug 2020 – Present

High Performance Computing Facilitator, CU Boulder & JILA

Boulder, CO

- Led 20+ large-scale C++/FORTRAN simulations producing 10+ TB of data; built end-to-end analysis pipelines identifying novel physics that overturned prior theoretical predictions.
- Processed large-scale HPC simulation outputs using MPI and GPU acceleration; applied DBSCAN/cKDTree clustering, a gradient-based optimization and signal processing to extract insights from high-dimensional data.
- Developed predictive models using gradient boosting to forecast system evolution from initial conditions; validated against simulation results to assess model reliability under extrapolation.
- Defined research questions, designed experiments, and pivoted approaches when results contradicted hypotheses, demonstrating independent decision-making when no established methodology existed.

Projects

Local RAG Chatbot for Technical Documents | Python, LangChain, ChromaDB, Ollama, Llama, FAISS

- Architected a privacy-first RAG system processing 400+ technical PDFs (2M+ tokens) with sub-second query latency, eliminating dependency on external APIs and ensuring complete data privacy.
- Implemented semantic chunking with 200-token overlap and embedding-based similarity search, achieving 85%+ relevance accuracy on domain-specific queries while maintaining full citation traceability to source passages.
- Designed modular 3-layer architecture (ingestion, retrieval, generation) enabling hot-swapping of embedding models and LLMs with zero pipeline restructuring, reducing iteration time for model comparison by 70%.

Black Friday Sales Prediction | Python, scikit-learn, XGBoost, Decision Tree, SHAP, pandas, datashader

- Built end-to-end ML pipeline benchmarking regression models on 550K retail transactions; XGBoost achieved best performance with 450% improvement over linear baseline in explaining purchases variance
- Deployed SHAP-based interpretability revealing Product Category and City Type explain 73% of prediction variance, and identified that EDA alone answered the core business question, avoiding unnecessary model complexity.

CCD Image Reduction & Solar Acoustic Analysis | Python, NumPy, SciPy, Astropy, FFT, SDO/HMI

- Developed 6-stage astronomical image pipeline (bias/dark/flat calibration, sub-pixel alignment, sky subtraction, RGB compositing) transforming raw CCD photons into publication-quality Ring Nebula portrait
- Implemented 3D FFT pipeline on 2 GB Solar Dopplergram datacubes ($512 \times 512 \times 1024$ voxels), successfully resolving solar p-mode ridges in Fourier space achieving SNR improvement of $12\times$ on raw Ring Nebula (M57) observations.
- Created fully reproducible workflow with documented parameters; pipeline reduces 4-filter raw data to publication-quality composite in < 5 minutes vs. 2+ hours manual processing.

Technical Skills

Tech Stack: Python (10+ yr), R, C++ & MATLAB (7+ yr), Bash, SQL, Git, HTML & TeX

Data Engineering & HPC systems: MPI, OpenMP, GPU parallelization, Frontier, Pleiades, Slurm

ML & Visualization: PyTorch, TensorFlow, LLM, Tableau, VisIt, Paraview

Leadership and Outreach

Founder and CEO, Entrepreneurship Cell :Built institute's first entrepreneurship program from scratch; led 8-member cross-functional team while balancing research, coursework, and teaching responsibilities.

Science Communication: Delivered astronomy education to 150+ public visitors at Fiske Planetarium and Sommers-Bausch Observatory; translated complex astrophysical concepts for non-expert audiences at conferences.