ARUNABH GHOSH

AI/ML Engineer

@ arunabhghosh98@gmail.com | 310-873-8845 | San Francisco Bay Area, California | arunabh98 | arunabh98

EXPERIENCE

Arista Networks
Data Scientist

April 2021 – Present | Santa Clara, USA

- Spearheaded development of an AI assistant leveraging **state-of-the-art LLMs** and **RAG architecture**, indexing over 200,000 technical documents, achieving **85%+ accuracy**; implemented **custom evaluation pipeline** using LLM-as-judge, A/B testing, and continuous feedback loops; **patent pending**.
- Authored **two patents**: "**Network Endpoint Entity Resolution**," improving device identification by consolidating various databases; and a **device risk scoring methodology** integrating vulnerabilities and threat metrics.
- Developed data anonymization algorithms using **NLP** and **NER**, detecting and **masking sensitive customer information**, enhancing data privacy compliance, enabling secure integration into AI products.
- Led **Device Hierarchy** integration into **Arista NDR**, standardizing device classification by unifying services like EDRs and other detection sources.

Chan Zuckerberg Biohub
Data Visualization Intern

June 2020 – September 2020 | San Francisco Bay Area, USA

- Developed a **glyph-based algorithm** to convert anisotropy and orientation data from polarization microscopy into **intuitive tensor visualizations**.
- Improved processing speed by **500%** using **parallelization**, **vectorization**, & precomputation; used **PyVista** color schemes for enhanced interpretation.

EPFL
Research Intern

June 2018 – July 2018 | Lausanne, Switzerland

- Implemented **Bayesian-optimized SGD** for ab-initio Cryo-EM protein reconstruction, outperforming established accuracy benchmarks.
- Applied **expectation–maximization** refinements, improving precision.

Google Summer of Code
Oppia Foundation, *7% acceptance rate*

May 2017 – August 2017 | Mumbai, India

- Engineered **full-stack learner dashboard** serving **400K+ users**, building high-throughput **MapReduce** pipeline for real-time tracking with personalized progress visualization, feedback management, and bookmarks.

PROJECTS

Molecular Structure Estimation, Thesis — *Prof. Ajit Rajwade, IIT-B*
- Developed a pipeline integrating **hierarchical clustering**, **graph-Laplacian** embeddings, and sparse optimization to reconstruct protein structures from noisy, outlier-prone data—no prior info required. **ICIP, 2019**

Efficient Training via Subset Selection — *Prof. B. Mirzasoleiman, UCLA*
- Implemented **gradient-based selection methods** to reduce CIFAR-10 training set by **40%** while maintaining accuracy through augmentation.

EEG-Based Motor Imagery Classification — *Prof. J. Kao, UCLA*
- Developed **CNN-LSTM** ensemble for 4-class motor imagery EEG classification, achieving **71%** accuracy through multi-subject optimization.

BERT-Based Toxicity Detection — *CS 263: NLP, UCLA*
- Built **multi-label BERT** classifier for toxic comments, achieving **72%** accuracy across 6 toxicity categories through advanced preprocessing.

Music Genre Classification with Random Forest — *CS 419: ML, IIT-B*
- Implemented **Random Forest** classifier with **Bayesian optimization**, achieving 56% accuracy across 9 music genres using novel timbre features.

EDUCATION

UCLA
Master of Science

2019 – 2021 | USA

- Electrical and Computer Engineering | GPA: **3.97**

IIT Bombay
Bachelor of Technology

2015 – 2019 | India

- Electrical Engineering (Hons) | GPA: **8.84/10**
- Minor Degree in Computer Science

PUBLICATIONS & PATENTS

- **Ghosh, A.**, "Network endpoint identification through network fingerprint based entity resolution," **U.S. Patent No. 11,799,882**, Oct. 24, 2023. *Assignee: Arista Networks Inc.*
- **Ghosh, A.**, "Device vulnerability risk assessment system," **Patent Pending**, Filed Sept. 2023. *Assignee: Arista Networks Inc.*
- **Ghosh, A.**, *et al.*, "Ab initio tomography with object heterogeneity and unknown viewing parameters," in **IEEE ICIP, 2019. (Oral)**

SKILLS

Programming & Tools
- **Languages:** Python (Primary), SQL, Scala
- **GenAI:** LangChain, LlamaIndex, Hugging Face, RAG systems, Weaviate (vector DB)
- **ML & Data:** PyTorch, TensorFlow, scikit-learn, NumPy, Pandas, BigQuery
- **Cloud & Deployment:** GCP (Vertex AI), Azure, Docker, Streamlit

Academic Courses
- **ML & AI:** Neural Networks & Deep Learning **(A+)**, Reinforcement Learning
- **Data Science:** Large-Scale Data Mining, Matrix Analysis **(A+)**, Big Data Systems
- **Image & Speech:** NLP, Image Processing & Computer Vision, Computational Imaging

Teaching Experience (UCLA)
- Mathematics for Life Scientists (UG)
- Feedback for Control Systems (Grad)
- Signals and Systems (Grad)
- Neural Networks & Deep Learning (Grad)

ACHIEVEMENTS

- All India Rank of **135** in JEE Advanced 2015
- **Silver Medal** at the 6th Inter IIT Technical Meet
- **Best Tech Prize**, Yahoo Hackathon, AR Game