

Jinam Shah

jinams47@gmail.com | github.com/jinamshah | linkedin.com/in/jinamshah

As a seasoned **Machine Learning Engineer**, I excel in developing advanced ML models and orchestrating large-scale pattern recognition on **250TB+** datasets, achieving **95% accuracy** in production. I specialize in architecting distributed systems that have achieved **70-80% improvements** in processing efficiency across cloud and HPC environments. My track record includes implementing **LLM-based** pipelines reducing inference costs by **60%**, engineering robust data platforms managing **900TB** data lakes, and leading cross-functional teams to deliver ML solutions saving organizations **\$1M+ annually**. I combine ML expertise with solid software engineering practices to build scalable, production-ready AI systems.

Work Experience

Software Engineer, Plant Sciences Initiative - North Carolina State University | July 2023 - Present

- Architected and implemented an end-to-end automated **drone imagery processing pipeline** handling **200GB+ per flight**, reducing research teams' operational overhead by **70%** through centralization
- Developed **HPC-based** image processing workflow using **Slurm/Singularity** for generating vegetation index datasets, achieving **\$100K+ annual savings** versus cloud alternatives
- Built scalable **full-stack** application (**Flask/Node.js**) with **multiprocessing** optimization achieving **sub-second** plot-level analytics on **5GB+** geospatial datasets
- Designed interactive **geospatial visualization** interface with custom plot-drawing tools, increasing researcher productivity by **60%** through automated field trial analysis

Machine Learning Intern, Cactus Communications Inc | May 2022 – Nov 2022

- Implemented and optimized **production ML pipeline** for researcher disambiguation using **AllenAI's S2AND**, achieving **95% accuracy** across **250TB+** of academic data
- Engineered critical performance improvements to S2AND's codebase, enabling large-scale processing, with contributions merged into the open-source project
- Developed **distributed ML architecture** using **AWS EMR** clusters for parallel model inference, reducing processing time by **60%**

Senior Software Engineer (Data Engineering), Cactus Communications Inc | June 2020 - July 2021

- Architected and led development of **large-scale ETL pipeline** processing **250TB+** academic data, achieving **80% improvement** in processing efficiency and storage management
- Designed distributed processing architecture utilizing **24K CPU cores** and **48TB RAM**, generating **4.5TB** of data in **2.5 hours** at **1/5th** proposed AWS cost
- Led **3 cross-functional teams** with **7 developers**, delivering **5 major releases** while reducing deployment failures by **90%**
- Spearheaded architecture planning for **ML/NLP products** processing **1M+ research papers** monthly, reducing inference costs by **60%**

Software Engineer, Cactus Communications Inc | June 2018 - May 2020

- Built **serverless architecture** handling **100K+ daily requests** with **sub-second** latency, achieving **99.99%** uptime through auto-scaling
- Optimized cloud infrastructure costs across **AWS/GCP/Azure**, reducing monthly spending by **40%** while improving performance by **3x**
- Developed and scaled **ML pipelines** processing **8TB** weekly data, reducing operational costs by **70%** through efficient resource utilization
- Led collaboration with **AWS S3 team**, resolving critical throughput bug affecting **500K+ transactions** daily, improving system stability by **85%**

Notable Projects

Scientific Document Classification GCP TPUs, TensorFlow, BERT, AWS ECS, Docker, S3, PySpark

- Engineered **BERT-based** ensemble achieving **2-min processing** (from 8hrs) for **1500-class** classification; saved **\$1M annually** via serverless architecture handling **100K+ daily** requests

Ethical Image Compliance System AWS Lambda, scikit-learn, OpenCV, S3

- Built serverless CV pipeline achieving **99.8% accuracy** across **50K+ monthly** image validations with **sub-second** latency

Explainable Grammar Correction PyTorch, Horovod, NLTK, spaCy, AWS EC2

- Developed **distributed LLM** training pipeline processing **100K+ papers monthly**; achieved **85% accuracy** in explanation generation

Bias Detection in Text PyTorch, transformers, NumPy, scikit-learn

- Achieved **92% bias reduction** while maintaining model performance within **3%** through novel vector normalization approach

Image Caption Generator PyTorch, CNN, LSTM, Flickr-30K

- Implemented attention-based adversarial network achieving **96% accuracy** on Flickr-30K dataset

Master's Thesis: Author Name Disambiguation PyTorch, transformers, Hugging Face, pandas

- Developed **LLM embedding** approach achieving **70% accuracy** without demographic features across **1M+ papers**

Education

Master of Science in Computer Science, North Carolina State University | GPA: 4.0

Coursework: High-Performance Machine Learning, Neural Networks, Natural Language Processing, Algorithms for Data-Guided Business Intelligence

Thesis: Author Name Disambiguation in Academic Records using Large Language Models

Bachelor of Technology in Computer Science, NMIMS University | GPA: 3.51

Coursework: Artificial Intelligence, Predictive Modeling, Data Warehousing and Management, Image Processing, Data Analytics, Software Engineering

Research: Image Recognition and Image Caption Generation

Technical Skills

- Machine Learning & AI:** PyTorch, TensorFlow, scikit-learn, BERT, Transformers, Neural Networks, LLMs, Computer Vision, NLP, Hugging Face, Deep Learning, Distributed Training, MLOps
- Big Data & Cloud:** Apache Spark, PySpark, AWS (EC2, S3, Lambda, ECS, EMR, Redshift, Kinesis, API Gateway), GCP (TPUs, BigQuery), Azure, Airflow, Data Lakes, ETL Pipelines, Pandas, NumPy, OpenCV, NLTK, spaCy, Geospatial Processing, Data Warehousing, Data Visualization
- Programming & Development:** Python, SQL, Node.js, Flask, Django, React.js, Git, REST APIs, Microservices, Docker, Singularity, HPC, Slurm
- Development Practices:** Distributed Systems, System Design, CI/CD, Serverless Architecture, Performance Optimization, Agile/Scrum, Technical Leadership