# Jinam Shah

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

Engineer with 5+ years of expertise designing and deploying scalable ML solutions in NLP, Computer Vision, and Big Data. Proven track record building robust, production-grade pipelines, optimizing distributed training workflows, and improving system performance through effective MLOps strategies. Skilled in leveraging cloud and HPC resources to deliver impactful AI systems. Adept at problem-solving, innovation, and translating research to practical applications.

# Work Experience

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

- Engineered a scalable data processing pipeline to process the largest agricultural imagery dataset (300K+ bot-captured images), automating transfer, storage, and annotation.
- Built plant detection model based on YOLOv11 for identifying crops, weeds, and calibration objects; customized bounding box handling for Ultralytics implementation.
- Automated iterative model retraining loop with YOLOv11 object detection, streamlining labeled data integration, enhancing annotation pipeline throughput by **4x**, and reducing manual efforts significantly through robust **MLOps** practices.
- Architected HPC-based image processing workflow (Slurm, Singularity), saving \$100K+ annually vs cloud.
- Developed an end-to-end drone imagery processing pipeline (handling 200GB+ per flight) with a full-stack geospatial analytics interface (Flask, Node.js), reducing research overhead by 70% and improving productivity.

## Machine Learning Intern, Cactus Communications Inc | May 2022 - Nov 2022

- Constructed and optimized production ML pipeline for researcher disambiguation using AllenAI's S2AND, achieving 95% accuracy across 250TB+ of academic data.
- Implemented distributed inference pipeline on AWS EMR, reducing processing time by 60%.
- Contributed critical scalability improvements to the open-source S2AND repo.

### Senior Software Engineer (Data & ML), Cactus Communications Inc | June 2018 - July 2021

- Engineered BERT-based (LLM) ensemble achieving 2-min processing (from 8hrs) for 1500-class classification. Built training and inference pipeline and interactive frontend; saved \$1M annually via serverless architecture handling 100K+ daily requests.
- Built serverless ML inference pipeline handling 100K+ daily requests with sub-second latency, achieving 99.99% uptime through auto-scaling.
- Deployed ML/NLP platform handling 1M+ papers/month with inference cost savings of 60%.
- Architected large-scale ETL pipelines and serverless ML infrastructure to process over 250TB of academic documents and 8TB of weekly data, achieving 80% improvement in throughput and 70% reduction in operational costs.
- Created a distributed data processing system using 24K CPU cores and 48TB RAM to generate 4.5TB in 2.5 hours at 1/5th the cost of proposed cost.
- Led collaboration with AWS S3 team to resolve critical throughput bug affecting 500K+ daily transactions, improving system stability by 85%.
- Optimized multi-cloud infrastructure (AWS/GCP/Azure), reducing spend by 40% and increasing performance 3x.

## **Notable Projects**

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

• Forumlated LLM embedding-based approach achieving 70% accuracy without demographic features across 1M+ papers.

#### Automated Plant Detection Pipeline | YOLOv11, PyTorch, Ultralytics, HPC, Slurm

• Trained object detection model for plant/weeds recognition using 300K+ annotated field images. Iteratively retrained model with new labeled data. Achieved high precision in field detection and reduced manual labeling efforts drastically.

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

• Realized 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.

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

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

#### 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.

### 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**

- ML & AI: PyTorch, TensorFlow, Keras, scikit-learn, XGBoost,, Transformers, BERT, Hugging Face, YOLOv11, CNN, RNN, LSTM, Attention Models, Neural Networks, Deep Learning, Computer Vision, NLP, Reinforcement Learning, Generative AI, Distributed Training, GPU Acceleration, TPU utilization, Model Explainability (XAI), Hyperparameter Optimization, Transfer Learning, MLOps
- Big Data & Cloud: Apache Spark, PySpark, Hadoop, Hive, AWS (EC2, S3, Lambda, ECS, EMR, Redshift, API Gateway, SageMaker), GCP (TPUs), Airflow, Kubernetes, ETL Pipelines, Data Lakes, Data Warehousing, Feature Engineering, Geospatial Analytics, NumPy, Pandas, OpenCV, NLTK, spaCy, Matplotlib, Seaborn, Data Visualization, SQL
- Programming & Development: Python, SQL, Flask, Django, React.js, JavaScript, REST APIs, GraphQL, Microservices, Docker, Kubernetes, Singularity, Git, Linux, Shell Scripting, HPC, Slurm, CI/CD, Jenkins, GitHub Actions
- Development Practices: Distributed Systems, System Design, Serverless Architecture, Cloud Architecture, Infrastructure as Code (IaC), Scalability, Performance Optimization, Agile Methodology, Scrum, Technical Leadership, Code Reviews, Continuous Integration & Deployment