Jinam Shah

Raleigh, NC | (919)922-8481| jbshah@ncsu.edu | github.com/jinamshah | linkedin.com/in/jinamshah

I am a software engineer with over three years of experience in the industry. I have developed and contributed to various applications in machine learning, natural language processing, and big data. I have experience working with all three major cloud providers - AWS, GCP, and Azure - and have successfully designed systems capable of handling large volumes of requests and optimized for cost-effective parallelization. I have also designed and implemented an NLP product capable of accurately classifying documents across 1500 classes. I have created a big data platform that can ingest and manage approximately 1.5 TB of data per day, totaling over 900 TB in the data lake.

WORK EXPERIENCE

Cactus Communcations Inc, USA
Machine Learning Intern *May 2022 – Nov 2022*

- Managed large-scale **pattern recognition** on **250 TB** of raw data.
- Served as a lead software engineer for a project focused on **disambiguating records** in a data lake, coordinating between 3 teams, 7 team members, and two external vendors.
- Designed and implemented a serverless, highly available API that efficiently handles **~100K long-running requests** with a **sub-second SLA**.

Cactus Communications, India
Senior Software Engineer *June 2020 – July 2021*

- Spearheaded the **architecture planning** and implementation of various ML/NLP and big data products, effectively **bridging the gap** between business and technical teams.
- Established and coordinated **open communication channels** with the machine learning teams at AWS and Azure.
- Designed and implemented a highly efficient **data processing pipeline** for a machine learning product, utilizing **24K CPU cores and 48Tb RAM** to generate over 4.5Tb of data in under 2.5 hours, all while achieving **1/5th the cost proposed by the AWS Big Data Team**.
- Developed a **big data platform** that ingests over 1.5TB daily and generates approximately 8TB weekly, managing a total of **900TB in the data lake**.
- Established best practices and **operational runbooks** for the team to effectively operate on cloud platforms, including AWS, GCP, and Azure.

Python Developer *June 2018 – June 2020*

- Identified and collaborated with the **AWS S3 team** to resolve a **prefix throughput bug**.
- Designed and developed **image recognition** products, leading the charge to create a **new business vertical** for the company.
- Successfully designed and implemented a scalable **ML product** from scratch in under a week, with **zero downtime**.

PROFESSIONAL PROJECTS

Transformer-based Document Classification

- Developed an ensemble of DL and ML models (based on **BERT**) for performing document classification across 1500 classes, utilizing a serverless architecture to deploy the system.
- **Saved** the organization approximately **$1M per year** and reduced the TAT for the service from **8 hours to under 2 minutes** (a 99.6% reduction).

Serverless Image Recognition

- Designed and implemented an image recognition software for determining the **ethical compliance of images** in research papers, achieving state-of-the-art performance (**99.8% accuracy**) through a **serverless architecture**.

Automated Language Correction

- Contributed to the development of an **explainable AI**-based NLP software for automated grammar correction, leading the efforts to build a **scalable infrastructure and API** for the product.

Bias detection in text

- Utilized NLP and statistical machine learning techniques to **identify stereotypical biases** in text, creating a tool for assessing data quality and integrating bias detection into the workflow for text-based decision-making companies.

Image caption generator

- Worked on the adversarial neural network for **text generation and image recognition**.
- Achieved **96% accuracy** on the Flickr-30 dataset.

EDUCATION

Master of Science in Computer Science, North Carolina State University *August 2021 – May 2023*
GPA: 4.0
Thesis: Academic author name disambiguation using research topics
Relevant Courses: High-performance Machine learning, Neural Networks, Natural Language Processing, Artificial Intelligence-1, Automated Learning and Data Analysis, Design and Analysis of Algorithms, and Software Engineering.

TECHNICAL SKILLS

Programming tools: Python, Pytorch, Tensorflow, Keras, Pandas, Django, Flask, Spark, SQL, Git, C, C++

Domain Expertise: Machine Learning, Deep Learning, Natural Language Processing, Image Recognition, Distributed Training, Big Data

Cloud technologies: AWS, GCP, AWS EC2, AWS S3, AWS Lambda, AWS RedShift, AWS Kinesis, AWS API Gateway, AWS CloudFormation, AWS ECS

Design Principles: Cost-effective, scalable, secure, reproducible, reliable, serverless architecture.

ACTIVITIES

Courses: Machine learning from Stanford University, AI programming with python nanodegree by Udacity.
Hackathons: Top 25 percentile in annual Reply Code hackathon with over 10000 teams worldwide, twice in a row.
Volunteer: AI4Good foundation, working on solving United Nations' Sustainable Development Goals.
Open source: Contributor of AllenAI's S2AND and Specter repositories, consistent contributor through HacktoberFest.