

ISHAQ AHMED SHAIK

U.S Work Authorized | 623-273-5788 | ishaq.sde@gmail.com | [linkedin.com/in/ishaq-sde](https://www.linkedin.com/in/ishaq-sde) | github.com/sonuishaq67

EDUCATION

Arizona State University

Master of Science in Computer Science • GPA: 3.8/4.00

- Relevant Coursework: Statistical Machine Learning, Artificial Intelligence, Data Processing at Scale, Cloud Computing

Tempe, AZ

Aug. 2024 – May 2026

Ramaiah Institute of Technology

Bachelor of Engineering in Information Science

Bangalore, India

Aug. 2019 – May 2023

TECHNICAL SKILLS

Languages Python, C++, Java, SQL, Go, JavaScript, TypeScript, Bash

ML & AI PyTorch, Deep Learning, Transformers, Reinforcement Learning, Computer Vision, RAG, Recommendation Systems

Data & Backend PostgreSQL, Redis, Kafka, DuckDB, Neo4j, Spring Boot, FastAPI, gRPC, Protobuf, Flask

Cloud & Infrastructure AWS (EC2, S3, Lambda, SQS), Kubernetes, Docker, SLURM, Linux, Terraform, CI/CD

Concepts Distributed Systems, ML Training Pipelines, Feature Engineering, Large-Scale System Design, Data Mining

EXPERIENCE

Graduate Student Assistant

Arizona State University

May 2025 – Present

Tempe, AZ

- Deployed a **Python** RAG classification pipeline serving 5,000+ users at 1.2s latency with AWS Lambda and schema validation.
- Reduced query error rate from **30%** to 1% by building automated retry logic and input validation in the ML serving pipeline.
- Cut multi-turn compute costs by **40%** by optimizing NLP payloads via context packing and memory summarization in Python.

Software Engineer

Extreme Networks

Jul. 2023 – Aug. 2024

Bangalore, India

- Achieved **99.99%** state recovery across 100K+ devices by architecting a **Java** Kafka pipeline for distributed offline syncs.
- Reduced inter-service latency by **25%** across 1M+ daily requests by migrating **Java** REST APIs to gRPC + Protobuf.
- Delivered 8 **Java** Spring Boot APIs for 20+ NFL/MLB stadium sites, validating 70+ edge cases for flagship hardware.
- Cut debugging time by **95%** (30 to 1 min) via a **Java** Spring Boot SSH API to pull lab AP logs and config diffs.

Software Engineer Intern

Nbyula

Mar. 2023 – May 2023

Bangalore, India

- Cut image load times by **84%** (2.5s to 400ms) via **JavaScript** responsive cropping and CloudFront edge caching at scale.
- Raised chat DAU by **32%** (3.2K to 4.2K) by fixing Redis cache invalidation bugs and optimizing **JavaScript** React rendering.

Software Developer Intern

LoadShare Networks

Sep. 2022 – Dec. 2022

Bangalore, India

- Eliminated manual **SQL** ops saving 15 min/day by building a **Java** Spring Boot Slack bot with automated query execution.
- Accelerated deployment by **75%** (40 to 10 min) by Dockerizing **Java** Spring Boot microservices with CI/CD automation.

PROJECTS

Chess Transformer Lichess Bot | PyTorch, ONNX, AWS EC2, python-chess | [sonuishaq67/chess](https://github.com/sonuishaq67/chess)

- Trained a **500M**-param transformer on 10B tokens using PyTorch mixed precision on **4xH100** GPUs with SLURM scheduling.
- Built an end-to-end **Python** data pipeline processing 100M+ games via DuckDB filtering, UCI tokenization, and mmap tensors.
- Deployed model to production with sub-second latency by exporting to ONNX and serving on **AWS EC2** via Lichess Bot API.

Pong Deep Q-Network Reinforcement Learning | PyTorch, Gymnasium, CNN | [sonuishaq67/pong](https://github.com/sonuishaq67/pong)

- Achieved **100%** win rate across 10K games by training a PyTorch Double DQN with experience replay and soft target updates.
- Accelerated training by **91%** by parallelizing RL across 16 AsyncVectorEnv workers on GPU cluster with torch.compile.
- Reduced observation space by **93%** via a **Python** grayscale, crop, and resize pipeline feeding 4×84×84 frames to a CNN.

Amazon Reviews RAG System | Python, PostgreSQL, pgvector, Neo4j

- Boosted precision@10 by **27%** over 571M reviews via a **Python** hybrid vector + knowledge graph retrieval pipeline.
- Identified **20%** precision gain from hybrid retrieval via **Python** ablation studies comparing vector-only and graph baselines.

ML Network Intrusion Detection | Scikit-learn, XGBoost, Pandas

- Achieved **99%** accuracy on NSL-KDD and CIC-IDS by training tuned XGBoost RF classifiers on **150K+** network flow samples.
- Reduced false positive rate by **40%** by engineering 12 traffic features and applying feature elimination across 41 raw attributes.

Distributed Social Media Backend | Go, MySQL, Redis, Kubernetes | [sonuishaq67/social-api](https://github.com/sonuishaq67/social-api)

- Built scalable **Go** backend sustaining 50K+ req/min on Kubernetes HPA, cutting query latency by 65% via Redis caching.
- Handled **10K+** concurrent users by implementing a Go gRPC service mesh and optimizing SQL query paths under load testing.