Ishaq Ahmed Shaik

U.S Work Authorized | 623-273-5788 | ishaq.sde@gmail.com | linkedin.com/in/ishaq-sde | github.com/sonuishaq67

Education

Arizona State University — Tempe, AZ
*Master of Science in Computer Science • GPA: **3.8/4.00*** — *Aug. 2024 – May 2026*
- *Relevant Coursework:* Cloud Computing, Data Processing at Scale, Distributed Systems, Statistical Machine Learning

Ramaiah Institute of Technology — Bangalore, India
Bachelor of Engineering in Information Science — *Aug. 2019 – May 2023*

Technical Skills

Languages C++, Python, Java, Go, JavaScript, TypeScript, SQL, Bash
Backend & Data Spring Boot, FastAPI, gRPC, Protobuf, Node.js, GraphQL, PostgreSQL, MySQL, Redis, Neo4j, DynamoDB
Cloud & Infrastructure AWS (EC2, S3, Lambda, SQS, CloudWatch), Kubernetes, Docker, Kafka, Linux, Terraform, CI/CD
ML & AI PyTorch, Neural Networks, Reinforcement Learning, Transformers, Generative AI, RAG, NLP, Computer Vision
Concepts Distributed Systems, Large-Scale System Design, Fault Tolerance, Scalability, Event-Driven Architecture

Experience

Graduate Student Assistant — May 2025 – Present
Arizona State University — *Tempe, AZ*
- Deployed a **Python** RAG-powered WhatsApp assistant serving **5,000+** users at 1.2s latency using AWS Lambda and SQS.
- Improved system reliability by reducing query error rate from 30% to **1%** via automated retries and schema validation.
- Reduced multi-turn compute costs by **40%** by optimizing **Python** NLP payloads via context packing and memory summarization.

Software Engineer — Jul. 2023 – Aug. 2024
Extreme Networks — *Bangalore, India*
- Achieved **99.99%** state recovery across 100K+ devices by architecting a **Java** Kafka streaming queue for offline syncs.
- Reduced inter-service latency by **25%** across 1M+ daily requests by migrating 12 **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 — Mar. 2023 – May 2023
Nbyula — *Bangalore, India*
- Cut image load times by **84%** (2.5s to 400ms) via **JavaScript** responsive cropping, lazy loading, and CloudFront edge caching.
- Raised chat DAU by **32%** (3.2K to 4.2K) by fixing Redis cache invalidation bugs and optimizing React rendering logic.

Software Developer Intern — Sep. 2022 – Dec. 2022
LoadShare Networks — *Bangalore, India*
- Accelerated deployment by **75%** (40 to 10 min) by Dockerizing **Java** Spring Boot microservices with CI/CD pipelines.
- Eliminated manual **SQL** ops saving 15 min/day by building a **Java** Spring Boot Slack bot with automated query execution.

Projects

Chess Transformer Lichess Bot | *PyTorch, ONNX, AWS EC2, python-chess* | sonuishaq67/chess
- Built an end-to-end **Python** data pipeline processing 100M+ games via DuckDB filtering, UCI tokenization + numpy memmaps.
- Achieved **0.3s** move latency serving live games by exporting the model to ONNX and deploying on **AWS EC2** with Lichess API.
- Trained a **500M**-param transformer on 10B tokens using PyTorch mixed precision on **4xH100** GPUs with SLURM scheduling.

Pong Deep Q-Network Reinforcement Learning | *PyTorch, Gymnasium, CNN* | sonuishaq67/pong
- Accelerated training by **91%** by parallelizing **Python** RL training across 16 AsyncVectorEnv workers with torch.compile.
- Achieved **100%** win rate across 10K games by training a PyTorch Double DQN with experience replay and soft target updates.
- Reduced observation space by **93%** via **Python** grayscale, crop, and resize pipeline feeding 4x84x84 stacked frames to a CNN.

Amazon Reviews RAG System | *Python, PostgreSQL, pgvector, Neo4j*
- Boosted precision@10 by **27%** over 571M reviews by designing 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.

Distributed Social Media Backend | *Go, MySQL, Redis, Kubernetes* | 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.

Open Source Contributions | *C++, Bash, Linux*
- Deskreen: Maintained Deskreen AUR package for 24 months, automating publish via **Bash** CI/CD triggered by release.
- xkb-switch: Fixed 3 broken flag combinations in xkb-switch by rewriting CLI parsing with getopt/optarg in **C++**.