Torab Shaikh

Senior Software Engineer | **AI Engineer**

hello@torabshaikh.com | github.com/torabshaikh | linkedin.com/in/torab-shaikh | torabshaikh.com

+1 310-870-9340 | Buffalo, NY (Open to Relocation)

Summary

Senior Software Engineer and AI Engineer with **6+ years of experience** spanning production AI systems, edge deployments, serverless cloud infrastructure, and full-stack development. Hands-on contributor to a production agentic AI framework with ∼1,000 daily downloads. Deployed real-time CV and LLM systems on **NVIDIA Jetson** hardware across 150+ devices in demanding field environments. Strong expertise in **Python**, **AWS**, **TypeScript**, and multi-modal AI. Recently completed M.S. in Robotics and AI at University at Buffalo.

Professional and Academic Experience

Graduate Research Assistant Jan 2025 – Jan 2026

Advanced Materials and Devices Lab, University at Buffalo *Buffalo, NY*

- No calibration baseline existed for the lab's 20x8 FSR sensor array; built a data acquisition and calibration pipeline in **Python** applying FFT-based signal processing and scikit-learn regression to predict forces up to 20N with **0.3N mean error**, meeting the target accuracy specification.
- Sensor data needed to reach concurrent ML classification and visualization consumers with no latency budget; architected a **C++** shared-memory pub/sub IPC library enabling low-latency inter-process communication without serialization overhead.

Tech Lead Jul 2021 – Jul 2024

Talent Litmus *IN*

- Inherited a 4-engineer cross-functional team with no defined workflows; established coding guidelines, structured code reviews, and sprint-based execution, reducing runtime issues by **90%** and technical debt by **70%**.
- Built Talent Litmus Genie, an LLM-powered product integrating **OpenAI GPT** and **Stability AI** APIs using **Python** and FastAPI, cutting content update turnaround by **80%**; also wrote Python automation scripts for deployment, status monitoring, and infrastructure health checks across 5 EC2 instances.
- Managed 120+ domains and SSL certificate lifecycle; migrated static assets to S3 and CloudFront; configured NGINX with Certbot to automate certificate rotation; built CI/CD pipelines cutting deployment time to **10 minutes** and scaling to **40,000 concurrent users**.
- Delivered 6 products with co-founders including mobile games with **50K+ Play Store installs**; contributed to a **120% increase** in client acquisition (58 to 128 clients) over 3 years.

Senior Software Engineer Jun 2020 – Jul 2021

Helios Web Services: Thinglogix Foundry Platform *IN*

- Clients sent batched multi-contract PDFs via email; built a **Python** pipeline on AWS Lambda that used OCR to read contract numbers, divided and arranged contracts accordingly, applied a custom **stamp and signature detection model** (OpenCV, scikit-learn) for verification, and uploaded organized records to Salesforce; reducing manual effort by **95%**.
- IoT-connected scanners captured ID documents requiring manual data entry; built a serverless pipeline that ran OCR via AWS Textract, classified document type using a trained **Python** classification model, and auto-filled corresponding Salesforce form fields; processing documents up to **500 pages**, 400% above the original requirement.
- Manual attendance tracking across a distributed workforce was error-prone; engineered a **Python/C++** multithreaded edge AI pipeline on **NVIDIA Jetson Nano** processing 3 simultaneous RTSP camera streams with face recognition, GPS integration, ONNX Runtime inference, and automatic reconnection logic; deployed across **150 buses** and 3-4 factories, reducing processing time by **99%**.

Software Engineer Jun 2018 – Jun 2020

Helios Web Services: Thinglogix Foundry Platform (AWS Special Partner, built by original AWS IoT team) *IN*

- Built and extended the Foundry multi-tenant IoT platform using AWS IoT Core, Lambda, API Gateway, SQS, DynamoDB, RDS, Cognito, IAM, Node.js, and TypeScript; enabled real-time device management and reduced custom per-client solutions by **80%**.
- Designed and shipped a developer SDK and Infrastructure-as-Code via CloudFormation templates covering dev, staging, and production environments; recognized with **"Excellent Technical Player of the Year"** award in 2019.

Projects

Griptape AI Oct 2023 – Aug 2025

Open Source Contribution *Remote*

- Contributed to Griptape, a **Python** agentic AI framework with ∼1,000 daily downloads; implemented a Redis-backed persistent memory driver for long-running agents and resolved a production bug; both changes merged into the main codebase.

AI Job Application Tracker | **tracker.cloudnode.tech** 2025 – Present
Personal Project *Remote*

- Built a live GenAI web application using **Python**, Google Gemini API, vector embeddings, React, and Firebase; features resume analysis, job posting parsing, semantic matching, Kanban board, GitHub-style activity graph, and a Chrome extension with multi-profile autofill; URL-based caching avoids redundant inference calls across users.

EdgeNarrator | **github.com/torabshaikh/EdgeNarrator** Feb 2026 – Present
Personal Project *Remote*

- Engineered a fully on-device multimodal inference pipeline on **NVIDIA Jetson Orin Nano** using **Python**; ingests live RTSP streams, runs Moondream2 VLM via Hugging Face Transformers, and streams narration to a web dashboard with browser TTS; benchmarked Ollama (∼1.5s/frame) vs Transformers (∼4s/frame); zero cloud dependency.

Skills

Programming Languages: Python, C/C++, JavaScript, TypeScript, Java, NodeJS
AI & Machine Learning: LLM Integration (OpenAI, Anthropic, Gemini), Agentic AI (Griptape, Ollama), PyTorch, TensorFlow, Keras, scikit-learn, CUDA, Computer Vision, Object Detection, ONNX Runtime, Hugging Face Transformers, Deep Learning, RAG, vector embeddings, VLM
Full Stack Frameworks: React, Angular, Next.js, FastAPI, Flask, NestJS, ExpressJS, RESTful APIs, HTML5, CSS3
Cloud Platforms: AWS (EC2, Lambda, ECS, S3, CloudFormation, API Gateway, SQS, SNS, RDS, DynamoDB, IoT Core, WorkMail, Textract, Rekognition, SageMaker, Bedrock, CloudFront, Cognito, IAM), Google Cloud (Firebase, Vertex AI, Gemini API)
DevOps & Infrastructure: Docker, Docker Swarm, Kubernetes, CI/CD, GitHub Actions, Bitbucket Pipelines, NGINX, Certbot, Bash, Shell Scripting
Databases: PostgreSQL, MongoDB, MySQL, DynamoDB, Redis
Robotics & IoT: NVIDIA Jetson (Nano, Orin Nano), ROS/ROS2, Gazebo, SLAM, Edge AI, MQTT, WebSockets, RTSP, GPS modules, Eclipse IOFog, sensor integration
Tools & Methods: Git, CMake, Linux, API Design, Agile/Scrum, Jira, Postman, uv, Claude Code

Education

University at Buffalo Buffalo, NY
Master of Science in Robotics and AI *Graduation: Dec 2025*

- Relevant Courses: Machine Learning, Deep Learning, Computer Vision, Reinforcement Learning, GPU Programming (CUDA), Robotics Algorithms, Control Systems for Robotics, Probability

Poornima University Jaipur, India
Bachelor of Technology in Computer Engineering *Graduation: May 2018*

- Relevant Courses: Operating System, DBMS, Data Structures and Algorithms, Digital Electronics, Microprocessor and Interfaces, Computer Networks, Artificial Intelligence, Neural Networks, Real Time Systems, Distributed Systems, Embedded Systems, Cloud Computing, Mobile Computing

Achievements

Chess – Four times Secured 1st position in different inter and intra-college chess competitions, 2014–2018.

Sudoku – Secured 1st Position in Himalaya Hostel Fest Sudoku Competition 2016.

PwC Deep Learning Challenge 2017 – Nation-wide winner (with 3 Person Squad); achieved 99.7% accuracy on MNIST using a Python deep learning model on an NVIDIA platform.

Excellent Technical Player of the Year, 2019 – Awarded by Helios Web Services for CloudFormation IaC delivery and developer SDK contributions to the Thinglogix Foundry platform.