# Romil Shah

San Jose, CA 95134
+1 8579197147 | shah.romil@northeastern.edu | rams16592@gmail.com | shahromil.com
github.com/shahromil16 | linkedin.com/in/shahromil16 | Google Scholar | Justia Patents

## PROFESSIONAL EXPERIENCE

**Amazon Web Services**, San Francisco, CA

*Senior Applied Scientist, AWS Elemental* — Feb 2025 - Present

- Leading scientific initiatives & technical direction for development of foundation models for multi-modal understanding in Project Starfish, driving innovation in video verticalization, highlight generation features; Increase video understanding accuracy by ~75%.
- Leading and owning the R&D for AWS Elemental's video processing solutions, pioneering AI-driven filters (compression, scaling, frame-rate conversion, transcoding) optimized for AWS Trainium/Inferentia chips, reducing inference cost by ~200%.

*Sr. Applied Scientist, AWS ProServe* — May 2023 – Feb 2025

- Led development of AI training & inference pipelines leveraging LLMs & VLMs to enable scalable AI solutions for AWS customers; Optimized multi-modal training pipelines using FSDP & FP8, and deployed edge-efficient models on humanoid robots for TRI.
- Developed and owned GenAI@Edge product offering that enabled over 10 customers to deploy CVML solutions at the edge; Successful deployments saved customers an average of $20M in automation costs and drove over $5M in average annual recurring revenue through adoption of AWS services.

*Applied Scientist II, AWS ProServe* — Aug 2021 – May 2023

- Architected & deployed DL pipelines on cloud & edge devices across industry verticals, enabling performant AI solution delivery.
- Led customer engagements to understand AI/ML needs and develop customized edge ML solutions, integrating learnings into a reusable product offering called EdgeML Accelerator; Developed MLOps best practices and workflows to enable continuous model improvement through retraining, redeployment, and monitoring.

**Dolby Laboratories**, Sunnyvale, CA

*Sr. Computer Vision & Imaging Engineer* — Dec 2019 – Aug 2021

- Led design, optimization, and deployment of computer vision and deep learning pipelines for Dolby ATG and Dolby iAPI products, enabling real-time video and image processing applications.
- Developed multimodal pipelines supporting segmentation, classification, object detection, GANs, and pose estimation to enhance Dolby's computer vision capabilities.
- Optimized model deployment on embedded devices, improving inference latency by nearly 400% to 25ms to support low-latency applications.

**Strada Labs**, San Francisco, CA

*Co-Founder and CTO (Part-Time)* — Dec 2018 – Dec 2019 *(Part-Time)*

- Co-founded an urban analytics startup applying computer vision to analyze city movements and assist urban planning and development, hereby automating the process & saving costs for city planning.
- Raised seed funding through Y-Combinator; developed optimized cycle lane solution integrated in Chinatown, San Francisco.

**Ford Research and Innovation Center, Ford Motor Company**, Palo Alto, CA

*AI Research Engineer* — Jan 2018 – Dec 2019

- Drove AI innovation for autonomous and connected vehicles.
- Enhanced CV solutions for Ford Performance Racing (NASCAR) using deep learning for pre-race, in-race, and post-race analysis.
- Published research applying CV, ML, DL and RL to mobility domains.

**Volvo Construction Equipment**, Shippensburg, PA

*Computer Vision Research Engineer Co-Op (8-month-internship)* — Jan 2017 – Aug 2017

- Developed and optimized object detection and tracking solutions on embedded systems for semi-autonomous construction vehicles, improving latency, detection accuracy.
- Integrated and fused camera, RADAR, and stereo vision data to enhance perception capabilities**.**

**ReGameVR Lab**, Boston, MA

*Research Assistant* — July 2016 – Dec 2016

- Rehabilitation oriented frontal/profile face detection using OpenCV libraries and Haar-like features using sensor fusion of camera system and IMUs; Using Kinect for tracking human body-joints to improve rehabilitation techniques and create a labelled dataset; UDP for IoT connection between Raspberry Pi and Arduino.

**Tellmate Helper Pvt. Ltd.**, Ahmedabad, India

*Chief Developer and Co-Founder* — May 2014 -Aug 2015

- Co-founded and led the making of 'Tellmate', a device made using Kinect360 and Intel RealSense camera integrated with PandaBoard ES for assisting visually impaired people; Selected for Top 20 startups in India from 1.9k participants; 200k INR seed funding by Intel Digital India Challenge 2015.

**Florida Atlantic University, Multimedia Lab**, Boca Raton, FL

*Summer Research Intern* — May 2013 – July 2013

- Video processing, scene analysis-characterization-clustering, compression using motion estimation and motion vectors; using X264 and FFMPEG.

## EDUCATION

**Northeastern University**, Boston, MA — Dec 2017
***Master of Science in Electrical and Computer Engineering***
*Concentration: Computer Vision, Machine Learning, Control Systems*

**Nirma University**, Ahmedabad, India — May 2014
***Bachelor of Technology in Electroncis and Communication Engineering***
*Concentration: Image Processing, Robotics*

## PATENTS AND PUBLICATIONS

- Shah, R., et al. 2025. Systems and Methods for Recommendation-based Multimedia Advertising using Generative Artificial Intelligence (GAI) Product Placement and feedback. U.S. Patent Application 18/757,243.
- Shah, R., et al. 2024. Systems and Methods for non-invasive beverage quality check and automated maintenance scheduler for beverage dispenser. U.S. Patent Application 18/141,315.
- Shah, R., et al. 2024. Systems and methods for tracking luggage in a vehicle. U.S. Patent Number 11/882,500. Granted Jan 2024.
- Shah, R., et al. 2023. Optimized recharging of electrical vehicles. U.S. Patent Number 11/609,571. Granted Mar 2023.
- Shah, R., et al. 2023. Vehicle damage identification and incident management systems and methods. U.S. Patent Number 11/562,570. Granted Jan 2023.
- Shah, R., et al. 2022. Vehicle yield decision. U.S. Patent Number 11/338,810. Granted May 2022.
- Shah, R., et al. 2019. Systems and Methods for seat selection in a vehicle of a ride service. U.S. Patent Number 11/170,459. Granted Nov 2021.
- Shah, R., et al. 2019. Systems and Methods of preventing removal of items from vehicles by improper parties. U.S. Patent Number 11/295,148. Granted April 2022.
- Shah, R., et al. 2019. Vehicle Damage Identification and Incident Management Systems and Methods. U.S. Patent Application 20220108115, filed October 2020.
- Shah, R., et al. 2019. Offline Proximity Rideshare Booking System. U.S. Patent Number 16/581,104. Granted Nov 2021.
- Shah, R., et al. 2019. Systems and Methods for tracking Luggage in a Vehicle. U.S. Patent Application 20220141621. Filed November 2020.
- Rivera, A., et al. 2018. Object Locator with Fiducial Marker. U.S. Patent Number 11/010,919, Granted May 2018.
- Balasubramanian, SN., et al. 2019. Ride Request Evaluation Systems and Methods. U.S. Patent Application 20200293953. Filed May 2019.
- McKenzie, M., et al. 2019. Optimized Recharging of Autonomous Vehicles. U.S. Patent Application 202102255633. Filed Feb 2020.
- Patel, S., Shah, R. 2013. Femtophotography for detection of microbends in step index fiber. IEEE INDICON'13, IIT Bombay, India

## PROFESSIONAL ACTIVITIES

**Certifications:**

- *AWS Certified AI Practitioner (AIF-C01) – Dec '24 to Dec '27*
- *AWS Certified Machine Learning - Specialty (MLS-C01) – Aug '24 to Aug '27*

**Public Talks:**

- *Warehouse automation with cutting-edge supply chain solutions – re:Invent, Dec '24*
- *GenAI at Edge – NVIDIA Jetson AI Lab Research, Jul '24*
- *Grounded content generation for Amazon advertisements using GenAI – AI21Labs, Dec '23*
- *Autonomous Driving Data Framework (ADDF) Workshop – re:Invent, Dec '23*
- *Optimization and deployment of AI models on edge framework – re:Invent, Dec '22*
- *Leveraging gaming to generate training data – AI DevWorld, Oct '19*

## SKILLS

- **Operating Systems**: Linux, Windows, Mac
- **Simulators**: MATLAB, Simulink, WireShark
- **Familiarity**: CUDA, LiDAR, PCL, ROS2, SLAM
- **Libraries**: OpenCV, ZeroMQ, Gstreamer, Dlib, NumPy, FFmpeg, TFJS, Node.js
- **Programming Languages**: Python, C++, Typescript, Javascript
- **Deep Learning Tools**: OpenVino, Caffe2, PyTorch, TensorFlow, AWS Services