Bhavya Minesh Shah

/623-332-9509 | /bshah43@asu.edu | /bhavya-minesh-shah | /BhavyaShah1234 | /search.asu.edu/profile/5016411

Summary

Robotics and AI engineer with an MS in Robotics and Autonomous Systems (AI) degree. Gained industry and research experience during and prior to his Master's degree. Has worked in different roles in industrial automation, healthcare and pharmaceutical rebate domains for over 2.5 years. Also possess experience in publishing research.

Experience

Robotics and AI Researcher, Courtesy Affiliate January 2026 – Present
Logos Robotics Lab at Arizona State University *Tempe, Arizona*

- Developed and evaluated robot learning policies including Action Chunking Transformers (ACT), diffusion models, and Vision-Language Action (VLA) on the data collected via teleoperation on a bi-manual robot.
- Created an annotation tool with automated annotations capability using dynamic finetuning and reinforcement learning of LLMs using the annotator's feedback.
- Benchmarked open-source ASR models and audio capable multimodal LLMs on EMS call project to decide if finetuning was required for off-the-shelf models based on Word Error Rate, BLEU and ROUGE score.
- Developed a Mujoco robot simulation for table-tennis and trained multiple reinforcement-learning policies (e.g., PPO, SAC).
- Second authored the following research paper: https://arxiv.org/abs/2603.19166

Software Engineer, Platform and Integrations May 2026 – August 2026
Anton Intelligence LLC *Tempe, Arizona*

- Migrated local file storage from an Azure VM to Azure Blob Storage, converting it to a PaaS solution and preventing data loss while ensuring compliance.
- Updated the Azure SQL Database writer in Python to capture and store metadata for each file uploaded to Azure Blob Storage, improving data traceability and simplifying audit processes.
- Created a config-driven report generation Python package to ease report generation by non-developers.
- Deployed a modular forecasting script to Azure Function App via Azure DevOps pipelines, providing on-demand forecast capabilities to the platform and improving user access to predictive insights.

Helios Scholar & Graduate Intern, AI Researcher May 2025 – May 2026
Translational Genomics Research Institute *Phoenix, Arizona*

- Designed and developed an Agentic AI platform for doctors and biomedical practitioners enabling HIPAA-compliant data handling, multimodal reasoning, and visual question answering (VQA) over patient imaging and clinical records.
- Built an LLM orchestration system using LangChain, LangGraph, and Ollama, connecting task-specific LLMs and VLMs (for de-identification, diagnosis, recovery tracking, and reasoning segmentation) under a single orchestrator LLM.
- Developed Reasoning and Referred Segmentation features to visualize LLM decision logic, improving transparency and interpretability for medical users.
- Researched LLMs and VLMs for AI-assisted Tumor Boards, training multimodal models with CLIP encoders on MRIs, CT scans, X-rays, and patient histories; evaluated performance using Rouge and BLEU scores on MedGemma, MedLlama3, Qwen3, etc.
- Optimized large-scale training on HPC clusters with SLURM, HuggingFace Transformers, Accelerate, PEFT, and BitsAndBytes; applied techniques like LoRA, quantization, and mixed precision for efficient multi-GPU training and evaluation.
- Incorporated Attention Pooling, Perceiver Resampling, and Unified Embedding Decoder using MediaPipe and ONNX, which streamlined multimodal processing and reduced inference latency; shared findings in a scientific poster and oral presentations at the institute's research symposium.

Artificial Intelligence Engineer August 2023 – May 2024
Accurate Industrial Controls Pvt Ltd *Pune, India*

- Built the AI division from scratch by setting up multi-GPU servers, CUDA/CuDNN environments, and GPU-compatible ML frameworks; established practices for scalable deployment.
- Developed and deployed perception and anomaly detection pipelines: finetuned YOLOv8 and WaSR (96.4% mAP@0.5) for autonomous boats, and optimized YOLOv8, PatchCore, EfficientAD, and EasyOCR with ONNX/TensorRT on Nvidia Triton for real-time defect detection.
- Designed the B-Star algorithm, boosting autonomous path-planning speed by 300% over A-Star and improved multi-sensor fusion reliability by introducing UNIX timestamp synchronization.

Publications

Meanings and Measurements: Multi-Agent Probabilistic Grounding for Vision-Language Navigation March 2026

Demonstrated Deep Learning Techniques for the Resolution of CAPTCHA images September 2022

Education

Master of Science in Robotics and Autonomous Systems (AI) August. 2024 – May 2026
Arizona State University (ASU), GPA: 3.57/4.0 *Tempe, Arizona*

Bachelor of Technology in Computer Science and Engineering (Bioinformatics) July. 2019 – April 2023
Vellore Institute of Technology (VIT) *Vellore, India*

Technical Skills

Programming Languages: Python, C/C++, JavaScript, Shell Scripting
Database Technologies: SQL, NoSQL, MongoDB, PostgreSQL, Redis
Frameworks/Packages: Django, Flask, FastAPI, LangChain, ROS Melodic and Noetic, ROS2 Humble, TensorFlow, PyTorch, Keras, MediaPipe, Selenium, Nvidia Triton Inference Server and Client, ONNX, TensorRT, HuggingFace Transformers, Accelerate, PEFT, Numpy, OpenCV, LangChain, Ollama, Nvidia Rapids, CuML.
Simulation/Modelling: Gazebo, MATLAB Simulink, RViz.
Developer Tools: Git, Github, CI/CD Pipelines, Jupyter Notebook, Google Colab, VS Code, Pycharm, Docker, Visual Studio.
Cloud Technologies: AWS, Microsoft Azure, SLURM, Google Cloud Platform (GCP)
Hardware: Nvidia GPU, Intel CPU, Raspberri Pi, Arduino Uno, ESP32, Nvidia Jetson Nano, Nvidia Jetson Orin, Intel NUC.
Agentic AI: Claude Code Terminal