

Devam Sheth

devamsheth21 | devamsheth21 | devamsheth20@gmail.com | 602.768.7274

EXPERIENCE

Software Development Engineer (SDE) <i>Amazon Web Services (AWS) – Bedrock Data Automation</i>	Mar. 2025 – Present Bellevue, Washington
<ul style="list-style-type: none">Led cross-functional coordination across 3+ teams to deliver customer-facing AI features on schedule, managing dependencies across distributed services and deployment pipelines.Optimized AI inference workloads through concurrency tuning, batching, and access pattern analysis, achieving 20% latency improvement and 30% cost reduction while processing millions of monthly requests.Led load testing and optimization of ECS-hosted Triton-based RAI service across 7 regions, identifying thread-bound bottlenecks and resource underutilization to achieve \$28K/month in cost savings through fleet rightsizing.Improved end-to-end request latency by 30% by redesigning client connection pooling and request concurrency mechanisms.Eliminated 45+ high-severity incidents by resolving systemic concurrency and race-condition issues in DynamoDB-backed workflows improving system stability under peak loads.Built comprehensive monitoring infrastructure with CloudWatch dashboards, custom metrics, and automated fault analysis scripts, reducing MTTD for production issues by 60%.	
Data Science Analyst <i>Mayo Clinic (AI & Informatics)</i>	June. 2024 – Mar. 2025 Phoenix, Arizona
<ul style="list-style-type: none">Developed Vision-Language Models to enhance diagnostic accuracy with multi-modal medical data.Built custom neural architectures using domain-specific pre-trained models and improved interpretability with 3D Grad-CAM.Engineered distributed training pipelines with custom data preprocessing for seamless multi-modal training.	
Machine Learning Research Assistant <i>Arizona State University, Wu Lab</i>	Aug. 2023 – May 2024 Tempe, Arizona
<ul style="list-style-type: none">Improved Brain MRI white matter segmentation with nn-Unet, achieving 0.65 mean dice score across 5 folds.Engineered distributed training pipelines and custom data preprocessing for seamless multi-modal training.Integrated LLMs and fine-tuned CLIP models, improving AUC by 8% for Headache classification and achieving 10% increase in classification performance on medical data.	
Machine Learning Engineer <i>Arrow Electronics Inc. (e-Infochips)</i>	May 2023 – Aug 2023 Ahmedabad, India
<ul style="list-style-type: none">Developed and deployed computer vision models for object detection, improving product quality by 15%.Optimized low-light enhancement models, reducing MACs by 82% through Neural Architecture Search.Optimized deep learning models for real-time inference on edge devices using distillation, quantization, and pruning.	
Machine Learning Research Assistant <i>Nirma University</i>	May 2021 – May 2022 Ahmedabad, India
<ul style="list-style-type: none">Developed bearing fault classification pipeline achieving 98% accuracy using EfficientNet CNNs with novel vibration-to-image feature engineering, FFT analysis, and transfer learning techniques.	

TECHNICAL SKILLS

Languages: Java, Python, C/C++, SQL (Postgres), JavaScript, TypeScript, HTML/CSS, R, Bash
ML Frameworks: PyTorch, HuggingFace, Llama-Index, LangChain, ONNX, Keras, TensorFlow, TensorRT, vLLM
Developer Tools: Git, AWS (Lambda, ECS, S3, DynamoDB), Docker, Kubernetes, Linux, Jira, Confluence
Libraries: Transformers, Diffusers, FAISS, OpenCV, Scikit-learn, Pillow, Seaborn, Pandas, NumPy, Matplotlib, OpenAI

EDUCATION

Arizona State University <i>Master of Science in Computer Science, GPA: 4.0/4.0</i>	Aug. 2022 – May 2024 Tempe, Arizona
Courses : Machine Learning, Natural Language Processing, Data Processing at Scale, Data Visualization	
Nirma University <i>Bachelor of Technology in Computer Engineering, Mechanical Engineering, GPA: 7.95/10.0</i>	July. 2018 – May 2022 Ahmedabad, India

PROJECTS

Hallucination Evaluation <i>LLMs, Hugging-Face, NLP, HPC Clusters, Git</i>	Oct 2023 – Sep 2023
<ul style="list-style-type: none">Led team evaluating MLLMs (InstructBLIP, Open-Flamingo) for hallucination detection using HPC clusters and developed end-to-end pipeline with CLIP-score filtering and Llama-based Q&A generation.	

ACHIEVEMENTS

Invited to join IEEE Eta Kappa Nu's ASU Chapter (Honors Society of IEEE) in Fall 2023
Graduate Fellowship Scholarship awarded in Fall 2023 by Arizona State University.