Sanika Chavan

+1 602 516 5693 • sanikac10@gmail.com • LinkedIn • GitHub

EDUCATION

M.S. Computer Science *May 2025*

Arizona State University, Tempe, AZ

PROFESSIONAL EXPERIENCE

Research Assistant, Arizona State University *June 2025*

- Implemented distributed training of large language models (**Qwen 2.5 Math 7B**) with CUDA, with **distributed training** (PyTorch, Flash Attention, vLLM,, TensorFlow) with **A100** GPUs, achieving a 2.8x training throughput increase through system optimization, **parallel processing**.
- Replicated and enhanced **Tiny Zero** reasoning framework using iterative **SFT** instead of **GRPO**, incorporating iterative fine-tuning with correct/incorrect labels (Reinforcement Learning) to improve model performance by 15%.
- Designed a comprehensive evaluation framework with automated metrics tracking, reducing assessment time by 70% while increasing result reliability.
- Engineered SLURM scripts for efficient **resource allocation** and **batch processing**, enabling 24/7 training utilization across shared compute infrastructure.

Software Developer Intern | *Vidyalankar Institute of Technology, Mumbai* *August 2021*

- Designed and implemented a Student Repository System using Java and MySQL with distributed architecture patterns, enabling efficient management of 10+ extracurricular activities across multiple department servers.
- Deployed a scalable system infrastructure with basic load balancing capabilities, facilitating institution-wide adoption within 3 months and supporting concurrent access from various administrative departments.
- Optimized database queries and implemented basic caching strategies, resulting in 2x reduction in administrative overhead and improved response times for student participation monitoring.

TECHNICAL PROJECTS

Orion - agentic cached workflow *August 2025*

- Orion breaks the déjà vu cycle by copying your workflow patterns and spinning off mini-agents so you never have to explain again. "Finally, an AI that remembers HOW you work, not just WHAT you said"
- Orion employs GEPA (Genetic-Pareto), a recent research breakthrough from Stanford and others that outperforms Group Relative Policy Optimization (GRPO) by 10-20% while using 35x fewer rollouts.
- This methodology evolves prompts through natural language reflection, what did we do with it? We made it into an automatic detection and caching agentic workflow.

Multi-Stage Reasoning Framework for VLLMs (Strong Compute Hackathon Winner) *April 2025*

- Engineered a synthetic data generation pipeline for the ARC-AGI-2 challenge benchmark, demonstrating superior general pattern recognition capabilities in our vision-LLM model.
- Structured model responses using three custom token blocks (visual explanation, logical planning, transformation code), which improved interpretability and enabled systematic reasoning on abstract visual patterns.
- Achieved 75+% resolution rate by training a reward model with SFT and GRPO, along with LIMO-based reasoning, significantly enhancing the model's ability to generalize across unseen pattern recognition tasks.

PUBLICATIONS

Data Leakage in LLMs (ICLR 2026) *August 2025*

- Designed and implemented a detection pipeline to identify key data leakage scenarios in large language models.
- Trained and evaluated five diverse (both open source and closed) models to analyze susceptibility to leakage, revealing key insights about model generalization, reasoning, and reliance on training data patterns.
- Developed and validated the Relevant Information module, leveraging techniques like Part of Speech tagging and entailment checks to detect data leakage and highlight the need for robust training processes and dynamic evaluation benchmarks.

Multi-modal Retrieval for Image, Tables and Text (WACV 2026) *July 2025*

- Crafted a multi-modal retrieval system which performs cross-embeddings on Images, Tables and Text to create object-object similarity between the different modalities.
- By improving the retrieval context by 20%, in process of creating a benchmark for information retrieval for cross-modality datasets like MMQA, compared and improved 7 competing baselines.

SKILLS

Programming Languages: Python, Swift, C++, Java, JavaScript, TypeScript, HTML, CSS, SQL

Tools and technologies: SwiftUI, Xcode, Unity Engine, Keras, MatPlotLib, Ski-kit Learn, Pandas, TensorFlow, MATLAB, Kafka, Apache Spark, PostgreSQL, NoSQL, Git, Jenkins, Docker, Hadoop, Distributed systems, REST APIs