

# Vihaan Patel

Tempe, AZ, USA • vihaan004@gmail.com • +1 (623) 320-6260 • <https://vihaanpatel.tech>

Computer Engineer specializing in rapid prototyping and deployment of AI/ML infrastructure. Experienced in translating complex agentic AI workflows, computer vision pipelines, and high-performance computing into scalable solutions.

## EDUCATION

---

### BSE Computer Systems Engineering (Honors)

Arizona State University, Tempe, AZ

May 2026

GPA 3.9

## WORK EXPERIENCE

---

### Research Intern (Arizona State University)

August 2025 – Present

- Built a domain-specialized agentic workflow achieving 96-100% coverage across sub-system level RTL designs, reducing token consumption 4x over baseline by encoding verification methodology into tools, prompts, and context retrieval.
- Used trace-level analysis to diagnose uncovered verification paths and inconsistent stimulus generation, identifying design comprehension as the bottleneck and proposing a workflow with improved context management and retrieval.

### Lead Tutor (Arizona State University)

January 2025 – May 2026

- Spearheaded the curriculum design and delivery of high-impact review sessions for CS, Math, and Physics, standardizing tutoring methodologies to ensure consistency across a network of 20,000+ students.
- Partnered with university faculty to execute campus-wide technical workshops, successfully scaling academic support operations and bridging the communication gap between course requirements and student success.

### Deep Learning Undergraduate Research Fellow (Arizona State University)

June 2024 – December 2024

- Developed scalable Computer Vision pipelines using PyTorch on HPC clusters to benchmark ConvNext and Detectron2 for complex pattern recognition tasks including classification, localization, and segmentation.
- Optimized Mask R-CNN architectures via custom data augmentations and supervised fine-tuning, successfully increasing model precision for automated defect identification and anomaly detection applications.

## RELEVANT PROJECTS

---

### Agentic Workspace for GPU acceleration

2025

- Built GPU-acceleration infrastructure utilizing NVIDIA RAPIDS (cuDF, cuML, CuPy) to standardize ETL data analytics pipelines, automate performance profiling, and deliver 3x speed-up on large-scale data workloads.
- Deployed a production-grade AI inference service on HPC clusters, utilizing robust state management and vector search (ChromaDB) to support complex query workflows; *NVIDIA AI Spark Challenge Awardee*.

### Unified AI Platform for College Students

2025

- Architected a centralized multi-agent platform integrating disparate university systems (registration, planning, career services) to resolve institutional data silos and provide a unified AI interface for data-heavy administrative workflows.
- Engineered a scalable Python backend using LangGraph for stateful agent orchestration, with persistent memory, dynamic tool execution, and Model Context Protocol integration to automate complex data retrieval and management.

## SKILL SET

---

- **Programming:** Python, C++, Java, TypeScript, Bash/Shell, SQL
- **GenAI:** LangChain, LangGraph, HuggingFace, OpenAI API, RAG, Vector Stores, MCP, vLLM
- **Machine Learning:** PyTorch, OpenCV, TensorFlow/Keras, Supervised Fine-tuning, Reinforcement Learning
- **Infrastructure:** CUDA, RAPIDS, HPC/SLURM, Docker, AWS EC2, GCP, Github Actions, Git
- **Web Development:** React.js, Next.js, Node.js, REST, FastAPI, OAuth, PostgreSQL