Vihaan Patel

Tempe, AZ, USA · vihaan004@gmail.com · +1 (623) 320-6260
vihaanpatel.tech · linkedin.com/in/vihaanpatel · github.com/Vihaan004

Publications

Understanding Inference-Time Token Allocation and Coverage Limits in Agentic Hardware Verification
Vihaan Patel, Aman Arora, Vidya A. Chhabria
8th ACM/IEEE International Symposium on Machine Learning for CAD, Jeju, South Korea, September 2026.
arXiv:2604.15657 · github.com/advent-lab/covagent

CovAgent: Agentic Hardware Verification
Undergraduate honors thesis, Barrett, The Honors College, Arizona State University, May 2026.

Research Experience

Research Intern - ASU Advent Lab, Arizona State University August 2025 – August 2026
Advisor: Dr. Aman Arora

- Built an agentic stimulus-generation workflow that reaches 96–100% coverage on sub-system-level RTL designs while consuming 4× fewer tokens than the baseline, by encoding verification methodology directly into the agent's tools, prompts, and retrieval rather than into a larger model.
- Studied where the additional inference budget stops paying off, and used trace-level analysis of uncovered paths to locate the bottleneck: the agent's comprehension of the design under test, not its stimulus-writing ability.

Deep Learning Undergraduate Research Fellow - Arizona State University June 2024 – December 2024

- Benchmarked ConvNeXt and Detectron2 transferred from natural-image pretraining to lung imaging across classification, localization, and segmentation, isolating where the domain shift cost the most accuracy; ran the sweeps on HPC/SLURM clusters with tooling that kept comparisons reproducible across seeds.
- Adapted Mask R-CNN to the target domain through custom augmentation and supervised fine-tuning, improving precision on anomaly and defect identification.

Projects

Agentic systems

PPAgent - RTL PPA optimization benchmark and agent Spring 2026

- Built a 5-task benchmark on picorv32 RISC-V core modules with a two-stage harness - Icarus Verilog for behavioral correctness, Yosys synthesis for PPA - scoring each solution on cell, flip-flop, and wire deltas against baseline. Extended mini-swe-agent and evaluated multiple LLMs against manually optimized reference solutions.

CovAgent - agentic verification workflow Spring 2026
See Research Experience.

MetaSolver - adaptive multi-technique reasoning agent Spring 2026

- Built a router for 7+ inference-time strategies - chain-of-thought, decomposition, tree-of-thought, self-consistency, tool-augmented arithmetic, self-verification and retry. Evaluated across 1,000+ questions with per-domain scoring.

Collagent - AI platform for college students Summer 2026

- Built a stateful orchestration with LangGraph - persistent memory, dynamic tool execution, MCP servers, and vector search - over data unified from disparate university systems, turning fragmented information into a queryable advisor.

GPU acceleration and CUDA

GPU-accelerated agentic data workspace Summer 2025

- Built ETL and profiling infrastructure on NVIDIA RAPIDS (cuDF, cuML, CuPy) that standardized data analytics pipelines, automated performance profiling, and delivered 3× speedup on large-scale workloads.
- Deployed an AI inference service on HPC clusters with robust state management and ChromaDB-backed retrieval to support complex query workflows. NVIDIA AI Spark Challenge awardee.

Custom CUDA kernels - PyTorch extension and ctypes bindings Fall 2025

- Wrote batched Linear-forward and ReLU-backward CUDA kernels and bound them two ways (PyTorch C++/CUDA extension, ctypes shared library), training MNIST end-to-end through both paths to compare framework integration against a from-scratch approach.

RTL design and verification

Systolic matrix-multiply accelerator - AMBA APB design + UVM verification Summer 2025

- Designed a 4×4 systolic accelerator in SystemVerilog with FP8 MAC units and an APB slave exposing a CSR map (operand addresses, strides, start, done); synthesized in Synopsys Design Compiler.
- Verified in UVM: an APB master VIP agent programs CSRs and polls done, custom memory-slave agents model the A/B/C SRAMs, and a virtual sequence drives a self-checking scoreboard with SVA and force/release fault injection.

Asynchronous FIFO - CDC design + DPI golden-model verification Summer 2025

- Designed a parameterized async FIFO with Gray-code pointers, dual-flop synchronizers, and almost-full/empty flags across independent clock domains; verified against a C/C++ golden model bound via SystemVerilog DPI, with plusarg bug injection confirming the checkers actually fire.

Pipelined MIPS processor - constrained-random stimulus + UVM functional coverage Summer 2025

- Generated legal instruction streams carrying RAW hazards, store-load dependencies, and branch outcomes, then closed functional coverage near 100% via covergroups merged across multi-seed runs with URG.

FPGA systolic array and multi-cycle MIPS extension Fall 2024

- Brought up a BRAM-backed systolic multiplier on Artix-7 and added a MULT16 instruction to a multi-cycle MIPS core, debugging RAM and PE failures through timing-driven simulation.

Additional projects: PAShell (POSIX shell in C with an LLM feedback layer) · Python Sandbox API (nsjail-isolated code execution on Cloud Run) · Pomodoro Max (Tauri desktop app, Rust + TypeScript) · autonomous MCU robot (bare-metal ARM firmware, PID control).

Education

Bachelor of Science in Engineering, Computer Systems Engineering August 2022 – May 2026

Arizona State University, Ira A. Fulton Schools of Engineering

GPA 3.87/4.00 · Summa Cum Laude · Barrett, The Honors College · Dean's List, 7 of 8 terms

Primary coursework: Foundations of Machine Learning · Natural Language Processing · Computer Architecture · Design and Synthesis of Digital Hardware · Advanced Digital Design and Verification · Embedded Microprocessor Systems · Operating Systems · Distributed Software Development · Applied Linear Algebra · Probability and Statistics · Image Processing and Analysis.

Teaching

Lead Tutor, Academic Support Network - 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.

Technical Skills

Hardware design and verification: SystemVerilog, UVM, SVA, constrained-random verification, DPI golden models, CDC design, RTL synthesis

EDA tools: Synopsys VCS, Verdi, Design Compiler, URG, QuestaSim, Vivado, Icarus Verilog, Yosys

Machine learning: PyTorch, transfer learning and supervised fine-tuning, reinforcement learning, scikit-learn, NumPy, OpenCV, TensorFlow/Keras

LLM and agent systems: LangChain, LangGraph, mini-swe-agent, MCP, RAG and vector stores, context engineering, tool-use and planning loops, LiteLLM, HuggingFace, OpenAI API

Evaluation and benchmarking: benchmark and harness design, per-domain scoring, multi-seed and checkpointed batch evaluation, functional coverage analysis

Systems and infrastructure: CUDA, vLLM, TensorRT, RAPIDS, HPC/SLURM, Docker, Linux, AWS EC2, GCP, GitHub Actions, CI/CD

Languages: Python, C++, C, Java, SQL, Bash