Mihir Patil

Computer Science at Purdue University | 2026
+1 (914) 261-2903 | mppat526@gmail.com | linkedin.com/in/0xMihir | github.com/0xMihir

Experience

Autopilot Compiler Intern June 2025 - August 2025
Tesla *Palo Alto, CA*

- Designed two **new MLIR dialects** to represent hardware features of next-generation neural network accelerator at the assembly and operation level. Wrote tests to verify IR legality for hardware constraints.
- Developed optimization and transformation passes to lower from PyTorch export to machine code, including register allocation, code generation and memory layout generation.
- Developed bridge between runtime infrastructure and RTL simulator to evaluate performance and numerical accuracy of compiled models, **greatly increasing** developer productivity without requiring physical hardware.
- Designed a compile-time memory allocator to efficiently manage tensors and data in SRAM using MLIR dataflow analysis.

Autopilot Silicon Validation Intern January 2025 - June 2025
Tesla *Palo Alto, CA*

- Developed bare metal GPU driver to manage MMU and job submission in RTOS, enabling rapid testing during SLT
- Designed tests to measure performance and functionality of GPU and UFS IP in hardware emulation for next-generation SoC
- Optimized OpenCL GEMM kernels on GPU, measured memory bandwidth between GPU, DDR, and accelerator blocks
- Developed bare metal driver for video decoder IP, allowing for performance measurement of video codecs during bring-up

Research Assistant August 2023 – Present
Purdue University *West Lafayette, IN*

- Safety-critical Simulation Generation for Autonomous Driving Systems
  - Designing unified model architecture with text-conditioned diffusion models to generate diverse and realistic driving scenarios for testing autonomous vehicle systems
  - Conducted experiments to evaluate lane and agent encoders for improved scene consistency and realism
- Archer: Hybrid and Adaptive Cold-Start Mitigation for Serverless on the Cloud and at the Edge
  - Implemented a genetic algorithm to optimize cloud servers and edge devices with keep-alive and prewarming policies, **reducing cold starts by 53%, execution time by 86% and energy overhead by 71%**
  - Developed serverless function orchestration framework for distributed edge devices, **reducing runtime by 60%**

Software Engineering Intern June 2024 - August 2024
Kudu Dynamics *Chantilly, VA*

- Developed an agentic LLM system with Microsoft Semantic Kernel that automatically identifies **10+** classes of vulnerabilities in commits, generates exploit trigger payloads, and patches bugs using delta debugging
- Contributed open-source fixes for Qdrant vector database support in the Semantic Kernel repository in PR #6950
- Created an automated software fuzzing system using LLMs to evaluate progress and evolve fuzzing corpus, increasing code coverage of open-source repositories in less time than standard fuzzers
- Developed LLM tool for generating binary data for exploit payloads from LLM plain text, increasing accuracy from **60% to 98%** over LLM binary output without tool calls
- Wrote proof of concept for exploiting the Linux Kernel using CVE-2024-1086, performing privilege escalation using nf_tables

Cloud Engineer March 2023 – December 2024
KeyByte *West Lafayette, IN*

- Led the development of Rust-based client software to collect database metrics with **zero overhead** using eBPF
- Built a self-hosted GitHub CI/CD pipeline, reducing developer time for unit testing from **hours to minutes** while increasing consistency and reproducibility
- Played a critical role in delivering an MVP over summer of 2024, designing robust system architecture and coordinating full-stack development and deployments
- Led the integration of MongoDB into a genetic algorithm to optimize database parameters and AWS instance size, allowing customers to increase performance by **up to 15%**

Education

Purdue University Aug. 2022 – May 2026
BS Computer Science Honors
Relevant Coursework: Machine Learning, Artificial Intelligence, Compilers, Systems Programming, Computer Architecture, Data Structures and Algorithms, Graphics, Linear Algebra

Projects

USCC Compiler | *LLVM, C++, x86 Assembly* September 2024 – December 2024

- Developed front end with lexer and recursive descent to convert C subset language an into abstract syntax tree
- Created LLVM optimization passes using data flow analysis for dead code elimination, LICM, and constant folding
- Implemented register allocation using Chaitin-Briggs graph coloring algorithm to minimize register spills

RescueVR | *PyTorch, Unity, OpenCV* January 2024

- Combined YOLOv8 predictions with structure-from-motion data to display distressed individuals in 3D space, achieving **98% accuracy** for 3D localization and action recognition
- Optimized inference time from **5 minutes to 30 seconds** per video, achieving near real-time video processing performance
- Developed LLM interface to enable natural-language scene understanding and visual question answering

Technical Skills

Languages: C/C++, Python, Rust, Typescript, Javascript, R, SQL, Java, Kotlin, Bash, HTML, CSS,
Technical Skills: MLIR, LLVM, Linux Kernel, Docker, Redis, Cassandra, PostgreSQL, InfluxDB, DevOps, eBPF, AWS, FreeRTOS, OpenGL, Reverse Engineering, Ghidra, PCB Design, KiCad, Altium, Fusion 360, Android Development, Raspberry Pi, STM32, Arduino, Svelte, Blender, Git, Unity