





ROBERT HOCHGRAF

Seeking Full-Time, AI Engineering, May 2025

 rhochgraf21
 robert-hochgraf
 rob@mail.rit.edu
 585-206-2372

EXPERIENCE

AI Engineering Intern

MITRE | June 2024 – Present

- Developed chatbot interface that automatically transforms company data into business knowledge
- Contributed to data retrieval systems, including by authoring PRs adding **agentic planning**, hybrid **graph-vector search**, and RAG optimizations (eg, reranking)
- Authored **prompt engineering copilot** to streamline prompt engineering process while also improving prompt performance by up to **50%**

LLMs RAG Llamaindex Neo4J Agents Azure

GPU Performance Researcher

University of Rochester | August 2022 – Jun 2024

- Achieved up to **7x speedups** over a state-of-the-art linear algebra library by formulating performance tuning framework for GPU kernels
- Presented findings at **Finger Lakes High Performance Computing Symposium** & University of Rochester seminars
- Authored supercomputing conference paper manuscript detailing the portability tuning technique (pending review)

Python C C++ Bash Linux OpenCL GPU

Software Engineering Intern

Chameleon Consulting | June 2023 – August 2023

- Researched and updated test-range infrastructure to support **secure deployments** onto production
- Architected intelligent onboarding automation script that slashed environment setup time from hours to minutes
- Authored internal exploratory research analyzing potential for machine-learning enabled capabilities

Go Docker Kubernetes TLS Proxies Git

AWARDS

Goldwater Scholarship Nominee

- For demonstrated achievement in STEM research

NYS Seal of Biliteracy

- For proficiency in both **Spanish** & **English**

Cameron Kent Buchanan Humanitarian Award

- For **compassionate altruism**

Audio Combat Champion

- For a remix that won the EDM Club's Audio Combat challenge

EDUCATION

Rochester Institute of Technology

BS, Computer Science, May 2025, GPA: 4.0

Concentration: Public Policy

SQL Pandas Cost Benefit Analysis ML Systems

PROJECTS

TritonTune

- Spearheading ongoing project to optimize performance tuning in the **Triton** tensor compiler for Machine Learning primitives.
- Designed performance model & identified optimization with promise to eliminate up to 1/3 of online tuning cost without impacting deployed performance.

Triton Python GPU Simulations Linux

Indirect App | iOS

- Industry-first app to train dentists' hand-eye coordination skills
- Ranked #25** in Education on iOS (1/28/2024)

Unity C# iOS Android Xcode

ServerlessCV

- Tensorflow** CNN model to classify content for moderation for the Computer Science House
- Containerized on **FaaS** solution & deployed entire solution on **OKD** instance

Kubernetes FaaS Tensorflow APIs Linux

ACTIVITIES

Computer Science House

Constitutional Maintainer | August 2022 – Present

- Author amendments to improve inclusivity, resolve constitutional crises, and enhance member engagement

RIT EDM Club

Music Producer | August 2022 – Present

- Avid Logic Pro user & producer of electronic music
- Scored music for student-led films in a variety of genres