

Neil McGlohon, Ph.D.

CONTACT

neil@mcglo.dev
nmcglo.com

github.com/nmcglo
linkedin.com/in/nmcglo

SKILLS

Software Engineering: C/C++, Python, JavaScript/Node.js, HTTP API design, AMQP, SQL, Git, CMake, Jenkins, ArgoCD
Parallel & Distributed Computing: MPI, SLURM, HPC systems, event-driven systems
AI/ML & Inference: PyTorch, TensorFlow, vLLM, AI agents, neuromorphic hardware
Cloud & Infrastructure: Kubernetes, OpenShift, Helm, Docker, Podman, Grafana, Prometheus, AWS, IBM Cloud, cloud-native systems, production-scale serving

PROFESSIONAL EXPERIENCE

– IBM, IBM Research

Almaden Research Center, San Jose, California USA

Staff Research Scientist

October 2022 – July 2026

- Led design and development of a purpose-built LLM inference engine for the IBM AIU NorthPole AI inference accelerator card, delivering state-of-the-art low latency and energy efficiency.
- Deployed rack-scale distributed AI systems of 288 NorthPole cards, generating over 500,000 tokens per second in aggregate while holding per-token latency as low as 1 millisecond.
- Led the team building the runtime stack and performance-critical SDK releases for NorthPole.
- Designed, built, and deployed a cloud-native OpenShift/Kubernetes stack that exposes and controls access to NorthPole-powered LLM inference appliances; integrated it into IBM watsonx, bringing NorthPole-accelerated LLM inference to workloads across the platform.
- Guided the team as subject-matter expert in parallel and distributed systems, containerization, orchestration, and LLM inference serving including vLLM integration.
- Represented NorthPole at business development events, client outreach and showcases, presenting the platform to prospective partners. Bridged scientific and business-focused teams, driving knowledge transfer and cross-team collaboration.

– Rensselaer Polytechnic Institute, Center for Computational Innovations

Troy, New York USA

Research Scientist

July 2021 – September 2022

- Led software development and optimization of workloads on the Center's HPC systems for internal and partner research objectives.
- Served on the steering committee of an NSF-funded project-of-projects, mentoring participating graduate students as a subject-matter expert in HPC and AI systems.
- Collaborated with the IBM AI Hardware Center on research into distributed AI training benchmarking and simulation of composable systems.
- Triaged and prioritized research-facilitation needs for thousands of users across multiple supercomputer systems.

Software Engineer

February 2020 – July 2021

- Maintained CODES, an open-source C/C++ framework for massively parallel simulation of high performance computing (HPC) communication networks, collaborating with researchers at Argonne and Sandia National Laboratories.
- Built custom software solutions in collaboration with external research partners.

– Cisco Meraki

San Francisco, California USA

Software Engineering Intern

May – August 2017

- Researched and developed firmware for Meraki's MR platform access points.
- Designed and built a lockless, thread-safe hashtable using RCU techniques, and upstreamed the contribution to Click, an open-source software-defined router project.

EDUCATION

- Ph.D., Computer Science** **July 2021**
 Rensselaer Polytechnic Institute, Troy, New York USA
 Research Focus: High Performance/Parallel Computing
- M.S., Computer Science** **May 2016**
 Rensselaer Polytechnic Institute, Troy, New York USA
 Research Focus: Distributed Computing, Machine Learning
- Bachelor of Science in Physics** **May 2014**
 University of Oklahoma, Norman, Oklahoma USA
 Minor: Computer Science

ACADEMIC
EXPERIENCE

- **Rensselaer Polytechnic Institute**, Troy, New York USA
Graduate Research Assistant **August 2014 – July 2021**
 M.S. and Ph.D. research, graduate-level coursework and projects. Areas of research: high performance and parallel computing, distributed systems, cloud computing, machine learning, neuromorphic computing and simulation.
Teaching Assistant: Computer Science I **August – December 2014**
 Supported the instructor of an introductory Python course: led lab sessions and discussions, held office hours, and proctored and graded exams/assignments.
- **University of Oklahoma**, Norman, Oklahoma USA
Undergraduate Research Assistant **May 2012 – May 2014**
 Explored electrical transport properties of quantum-confined antimony measured at cryogenic temperatures. Created and refined laboratory procedures for measuring differential conductance of a material, allowing for a deeper understanding of the interface between a topological insulator and a superconductor.
National Science Foundation REU **May – August 2012, 2013**
 Condensed matter research within NSF Research Experience for Undergraduates program including first observations of weak anti-localization in quantum-confined antimony films.

SCIENTIFIC
SERVICE

- NSF CAREERS Cyberteam Program
 Steering Committee Member (2020–2022)
- ACM SIGSIM Principles of Advanced Discrete Simulation (PADS)
 Organizing Committee, Ph.D. Colloquium Chair (2022)
 Organizing Committee, Publicity Chair (2023)
 Program Committee Member (2022–2025)
- ACM SIGPLAN Benchmarking in the Data Center: Expanding to the Cloud (BID)
 Program Committee Member (2022)
- IEEE IPDPS Scalable Deep Learning over Parallel and Distributed Infrastructure (ScaDL)
 Program Committee Member (2022)
- Peer-reviewer for 6 ACM/IEEE computing conferences and journals. (2019–Present)

PUBLICATIONS

Published 18 papers in conferences and journals spanning AI/ML, neuromorphic computing, high performance computing, and parallel discrete event simulation. Full list at nmcglo.com/cv.