

Neil McGlohon, Ph.D.

CONTACT

neil@mcglo.dev

nmcglo.com
github.com/nmcglo

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

RELEVANT SKILLS

- Self-motivated, fast-learner, quick to ramp-up on new topics
- Strong and flexible critical-thinking, problem solving, and analysis skills
- Expertise in parallel, distributed, and clustered computing
- C/C++, Python, Javascript/NodeJS, MPI, Git, SLURM, Cloud-native architectures, AI inference, HTTP API design, AMQP, vLLM, Tensorflow, PyTorch, Kubernetes, OpenShift, Docker and Podman containerization, AWS, IBM Cloud, L^AT_EX

FEATURED PROFESSIONAL EXPERIENCE

- Led the design and development of a purpose-built LLM inference engine to provide generative AI inference workloads powered by IBM AIU NorthPole hardware to define the state-of-the-art.
- Demonstrated rack-scale distributed AI system deployments of 288 NorthPole cards, capable of collectively producing over 500,000 generated tokens-per-second whilst simultaneously maintaining model-optimal individual per-token latencies as low as 1 millisecond.
- Designed, developed, and deployed a cloud-native stack on OpenShift/Kubernetes to expose and control access to IBM AIU NorthPole powered LLM AI inference appliances. Worked with engineers to integrate the services into the IBM watsonx product to enable IBM AIU NorthPole-accelerated LLM inference workloads across the platform.
- Acted as primary maintainer of open-source C/C++ project for massively parallel simulation of high performance computing (HPC) communication networks known as CODES. Collaborated with researchers from institutions including Argonne and Sandia National Laboratories.
- Participated as a steering-committee member on an NSF funded project-of-projects and acted as a subject-matter-expert and mentor in the areas of high performance computing and AI systems for participating graduate students.
- Author on 18 papers published in various conferences and journals in areas of AI/ML, neuromorphic computing, high performance computing, and parallel discrete event simulation.

EMPLOYMENT EXPERIENCE

- **IBM**, IBM Research
Almaden Research Center, San Jose, California USA

Staff Research Scientist

October 2022 – Present

- Team lead role driving and delegating tasks for the the development of the IBM AIU NorthPole accelerator card's runtime software stack and SDK packaging/release process.
- Subject matter expert in parallel and distributed computing systems, containerization and orchestration, and the serving of LLM inference systems.
- Technical representative for IBM AIU NorthPole during business development events and client outreach meetings as well as presentation showcases.
- Facilitated knowledge transfer, collaboration, and integration between teams across domains in both scientific and business-focused areas.

- **Rensselaer Polytechnic Institute**, Center for Computational Innovations
Troy, New York USA

Research Scientist**July 2021 – September 2022**

- Contributor to complex scientific efforts at the Center for Computational Innovations, focusing on software development and benchmark optimization in supercomputing environments.
- Identified and pursued funding opportunities to support innovative research projects in collaboration with the Center Director.
- External collaborator on research projects for the IBM AI Hardware Center.
- Provided technical expertise to enhance design, implementation, and analysis of research.
- Triaged and prioritized center needs for research facilitation of thousands of users across multiple supercomputer systems.

Software Engineer**February 2020 – July 2021**

- Development of research software for simulating high performance computing networks.
- Collaborated with external research partners to create purpose-built software solutions.

- **Cisco Meraki**

San Francisco, California USA

Software Engineering Intern**May – August 2017**

- Research and development of firmware for Meraki's MR Platform access points.
- Designed and developed a lockless thread-safe hashtable utilizing RCU techniques. Contributions submitted to open-source software-defined router project: Click.

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)

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, neuro-morphic computing and simulation.

Teaching Assistant: Computer Science I**August – December 2014**

Support to professor during course teaching fundamentals of computer science using Python. Facilitated laboratory classes and discussions, hosted office hours, graded homework and exams. Worked in group of eight teaching assistants.

- **University of Oklahoma**, Norman, Oklahoma USA

Undergraduate Research Assistant**May 2012 – May 2014**

Explored electrical transport properties of 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. Advisor: Dr. Sheena Murphy.

National Science Foundation REU**May – August 2012, 2013**

Participated in condensed matter research and analysis under an NSF Materials Research Science and Engineering Center (MRSEC) grant as part of an NSF Research Experience for Undergraduates program.