

Aly Sultan

978-325-1925 | Boston, MA

sultan.a@northeastern.edu | [linkedin.com/in/aly-sultan/](https://www.linkedin.com/in/aly-sultan/) | github.com/asultan123

EDUCATION

Northeastern University <i>Ph.D. Computer Engineering</i>	Boston, MA <i>Expected 2025</i>
Northeastern University <i>M.S. Electrical and Computer Engineering</i>	Boston, MA <i>2023</i>
American University in Cairo <i>B.S. Electronics & Communication Engineering</i>	Cairo, Egypt <i>2019</i>

EXPERIENCE

Graduate Software Engineering Intern <i>System Simulation and Modeling Group, Intel</i> <i>AI Cost Reduction in Simics SWCI</i>	2022 – Present <i>Part Time, Remote</i>
<ul style="list-style-type: none">Developed an AI solution to predict regression test failures based on developer source changes committed to GitAI solution aims to save computational resources by running test subsets more likely to failEstablished a developer metadata collection pipeline managing up to 2000 builds per week across 2 Simics platforms using Jenkins, Splunk and GitHub's GarphQl APIValidated collected data using JSON schema and produced daily data health-check reportsTrained XGBoost model on metadata collected and achieved up to a 40% reduction in regression test compute time on Granite Rapids and Diamond Rapids Simics models with a miss rate of 5.35% for failing testsProductized AI solution by creating a Pretest prediction tool piloted by 5 developers in the SSM Server teamShared project insights at Intel's internal AI Everywhere Conference and the S3E Tech Exchange	
<i>Secure Software Services Module (S3M) Firmware Integration Pipeline</i> <ul style="list-style-type: none">Established S3M's firmware integration pipeline, achieving daily FW deliveries for S3M's Simics modelCreated a versatile Python shell library, streamlining local and remote build operations across geographically dispersed data centers	
Graduate Research Assistant <i>Embedded Systems Lab, Northeastern University</i> <i>Hybrid General Matrix Multiplication and Direct Convolution Architecture (HERO)</i>	2020 – Present <i>Boston, MA</i>
<ul style="list-style-type: none">Developed a SystemC model for HERO, a novel matrix multiplication and convolution accelerator for DNN inferenceIntroduced Self Addressable Memory (SAM) for adaptive on-chip data orchestration in HEROEstablished HERO-SIM, a PyTorch-SystemC based simulation framework for the HERO acceleratorEvaluated HERO's efficacy on 695 DNNs, achieving up to 30X speedup and 300X energy savings over a workstation-class CPUSubmitted HERO manuscript to DAC 2024	
<i>Categorized Ensemble Networks for Adversarial Attack Defence (CAEN)</i> <ul style="list-style-type: none">Lead an AI defense project focused on bolstering ensemble network resilience against image-based adversarial attacksDeveloped a novel training methodology combining soft labeling with dissimilar label pairing, formulated the problem as an ILP, and solved it with GurobiTraining methodology achieved a 1.1X increase in robust accuracy over SOTA while reducing FLOPs by 16.8%Submission of CAEN manuscript to SPIE's DCS24 conference pending	

TECHNICAL SKILLS

Languages: *Python, C/C++, SystemC*
Framework: *Intel Pin, Darknet*
Developer Tools: *Git, Docker, Jenkins, QEMU, Simics, Gurobi*
Libraries: *PyTorch, Numpy, Pyomo*