Job Posting:173175 - Position: S26 2026 ML Compiler Software Engineer PEY Co-op (12-16 months), AWS Neuron, Annapurna Labs 173175

Co-op Work Term Posted:	2026 - Summer
App Deadline	09/19/2025 09:00 AM
Application Method:	Through Employer Website
Posting Goes Live:	09/12/2025 09:55 AM
Job Posting Status:	Approved

ORGANIZATION INFORMATION

Organization	Amazon.com
Country	Canada

JOB POSTING INFORMATION

Placement Term	2026 - Summer
<b> Job Title <b>	S26 2026 ML Compiler Software Engineer PEY Co-op (12-16 months), AWS Neuron, Annapurna Labs 173175
Position Type	Co-op Position
Job Location	Toronto, ON
Country	Canada
Duration	12 or 16 months
Work Mode	To be confirmed
Salary Currency	CAD
Salary	Salary Not Available, 0 Major List

Job Description

DESCRIPTION

At AWS, our mission is to make deep learning accessible to every developer by democratizing access to cutting-edge infrastructure. To achieve this, we've built custom silicon (AWS Inferentia and Trainium) and the AWS Neuron SDK that together deliver high-performance, cost-effective machine learning in the cloud.

The AWS Neuron SDK includes a compiler, runtime, debugger, and libraries integrated with popular frameworks such as PyTorch and TensorFlow. It is preinstalled in AWS Deep Learning AMIs and Containers so customers can quickly get started with training and inference on AWS ML accelerators.

The Neuron Toronto team focuses on performance, kernels, and tooling-analyzing and optimizing end-to-end ML workloads, developing and maintaining highly optimized kernels, and building performance modeling, profiling, and accuracy debugging tools. Together, these efforts ensure that Neuron delivers best-in-class performance, flexibility, and usability for customers deploying large-scale machine learning models.

As a student intern, you will contribute to the efforts that make Neuron best-in-class for ML workloads. You'll gain hands-on experience building business-critical features, analyzing performance, developing compiler or kernel optimizations, and building tools that provide deep insights into model execution. You'll be mentored by experienced engineers while working on technology that directly accelerates customer workloads at scale.

Job Requirements

BASIC QUALIFICATIONS

- Are enrolled in a academic program that is physically located in Canada
- Are enrolled in a Bachelor's degree or higher in Computer Science, Engineering Science, Computer Engineering, Electrical Engineering, or majors relating to these fields with an anticipated graduation date between May 2027 - May 2028
- Strong interests and academic qualifications/research focus in two of the following: 1. Knowledge of code generation, compute graph optimization, resource scheduling 2. Compiler - Optimizing compilers (internals of LLVM, clang, etc) 3. Machine Learning frameworks (PyTorch, JAX) and Machine Learning (Experience with XLA, TVM, MLIR, LLVM) 4. Kernel development-experience writing CUDA kernels, OpenCL kernels, or ML-specific kernels
- Available for a 12-16-month internship starting May 2026

PREFERRED QUALIFICATIONS

- Previous technical internship(s) related to the areas of interest / research focus listed above
- Experience in optimization mathematics such as linear programming and nonlinear optimization
- Academic coursework in Compiler Design/Construction, Programming Language Theory, Computer Architecture, Advanced Algorithms & Data Structures

Citizenship Requirement N/A

APPLICATION INFORMATION

Application Procedure Through Employer Website

Special Application Instructions

Please click the "I intend to apply to this position" button on SCOPE and also submit your application via the employer's website.
Application Link: 2026 ML Compiler Software Engineer PEY Co-op (12-16 months), AWS Neuron, Annapurna Labs - Job ID: 3081240 | Amazon.jobs
Applications are accepted on a rolling basis and the posting may be expired at any time by the employer as submissions are received.
Students should submit their applications as soon as they are ready.