

Job Posting:168129 - Position: S25 ML Compiler Software Engineer PEY Co-op (12-16 months), Annapurna ML 168129

| | |
|--------------------------------|--------------------------|
| Co-op Work Term Posted: | 2025 - Summer |
| App Deadline | 04/07/2025 09:00 AM |
| Application Method: | Through Employer Website |
| Posting Goes Live: | 03/24/2025 12:50 PM |
| Job Posting Status: | Approved |

ORGANIZATION INFORMATION

| | |
|---------------------|------------|
| Organization | Amazon.com |
| Country | Canada |

JOB POSTING INFORMATION

| | |
|--------------------------------------|---|
| Placement Term | 2025 - Summer |
| Job Title | S25 ML Compiler Software Engineer PEY Co-op (12-16 months), Annapurna ML 168129 |
| 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

At AWS our vision is to make deep learning pervasive for everyday developers and to democratize access to cutting edge infrastructure. In order to deliver on that vision, we've created innovative software and hardware solutions that make it possible.

AWS Neuron is the SDK that optimizes the performance of complex neural net models executed on AWS Inferentia and Trainium, our custom chips designed to accelerate deep-learning workloads.

The Neuron SDK consists of a compiler, run-time, and debugger, integrated with Tensorflow, PyTorch, and MXNet. It is preinstalled in AWS Deep Learning AMIs and Deep Learning Containers for customers to quickly get started with running high performance and cost-effective inference and training.

The Neuron team is looking for students interested in learning about and contributing to deep learning compilers to solve our customers toughest problems. As a compiler engineer on the Neuron team, you will be supporting the development of a compiler targeting AWS Inferentia and Trainium - our custom Silicon ML accelerators. This is an opportunity to work on a cutting-edge product where you will implement business-critical features, and be mentored by a brilliant team of experienced engineers.

Job Requirements

BASIC QUALIFICATIONS

- Experience with at least one modern language such as Java, Python, C++, or C# including object-oriented design
- Currently enrolled in a Bachelor's degree program or higher majoring in Computer Science, Computer Engineering, Electrical

Engineering or related technical field with anticipated graduation date on or later than May 2026

- Candidates with strong interests and academic qualifications/research focus in two of the following: 1. Knowledge of code generation, compute graph optimization, resource scheduling, 2. Data structure and algorithms, 3. Compiler - Optimizing compilers (internals of LLVM, clang, etc), 4. Machine Learning - Experience with XLA, TVM, MLIR, LLVM, 5. Machine Learning frameworks (PyTorch, JAX) and Machine Learning models
- Available for 12-month internship starting May 2025

PREFERRED QUALIFICATIONS

- Previous technical internship(s)
- Knowledge code generation, compute graph optimization, resource scheduling
- Experience in optimization mathematics such as linear programming and nonlinear optimization

Citizenship Requirement N/A

APPLICATION INFORMATION

Application Procedure Through Employer Website

Cover Letter Required? Optional

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: **ML Compiler Software Engineer PEY Co-op (12-16 months), Annapurna ML**

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.