Embedded AI developer for DNN runtime

About the Job

Are you a passionate software engineer with interest in Embedded systems (IoT) and Deep Neural Networks (DNN), seeking a thrilling opportunity to become a key technical contributor at an innovative AI startup backed by the European Space Agency (ESA)?

Join us at NinjaLABO, where we're developing DNN runtimes for IoT devices, to be integrated into our TinyML as-a-Service (TinyMLaaS) platform (research & demonstration). Your work will directly impact the accessibility of TinyML technologies, which run small AI on microcontrollers in IoT devices.

What We Need to See

1. Good at C (or C++) programming.
2. Familiar with training DNN models using Pytorch (or TensorFlow).

What You'll Be Doing

Have a role in the development of DNN runtimes, focusing on hands-on code contributions with DevOps.

- Train DNN models with quantization techniques.
- Implement DNN **runtime**s for specific DNN models

Bonus Skills

OpenMP, TensorFlow Lite For Micro (TFLM), Llama.cpp / Whisper.cpp / GGML / Llama2.c, CUDA, MLIR, Apache TVM, Mojo

Your Work Environment

- Fully remote position utilizing SCRUM methodology.
- Utilization of Docker and Docker Compose.
- Workflow managed through GitHub Projects, GitHub Workflow, GitHub Pages (Quarto),

Contract Period

6 months from March to August / 3 months from June to August

Monthly Pay

Contracts facilitated through UKKO.fi.

- 3,000+€ PhD student

- 2,600-2,800€ Master student
- 2,100-2,400€ Bachelor student

How to Apply

Interested in this groundbreaking opportunity? Direct your application to Hiroshi Doyu <hiroshi.doyu@ninjalabo.ai>