

AI Inference Software Developer

Flex Logix has finished the hardware design and is fabricating its first Inference Accelerator Co-Processor, InferX X1, which is based on our nnMAX Inference IP. We will have chips and PCIe boards this summer. Our software team is preparing our Inference Model Compiler to be ready to run deep neural network models on X1.

We have begun architecting the follow-on chip. InferX has industry-best inference efficiency: more inference throughput per \$ and per watt. We excel on larger models and megapixel images, but can run any neural network.

RESPONSIBILITIES

Part of the small but excellent team responsible for our nnMAX Model Compiler: a DNN Model-to-binary flow:

- Expanding functionality of our Model Compiler, written in modern C++, for support of additional capabilities, in particular:
- Parsing of TensorflowLite/ONNX/other DNN model description languages to our internal model format
- Mapping of required computations from DNN model in TFLite/ONNX to Verilog RTL code, running on EFLX eFPGA inside nnMAX chip.

This is a software developer role but you need to understand computer architecture and digital logic, as Model Compiler produces Verilog code for eFPGA, which controls nnMAX computational blocks and memory connections.

Read more about nnMAX and InferX on our Inference page at www.flex-logix.com

EXPERIENCE AND SKILL REQUIRED

BSCS/MSCS/BSEE/MSEE with courses in software/computer architecture/digital logic & 3+ years of relevant industry experience. AI/NN expertise is not required but experience is preferred in TensorFlowLite, Multi-core programming, and/or Windows and Linux, C++ and FPGA synthesis tools such as Synopsys Synplify.

Understanding of software/hardware development engineering practices, people with previous internships and/or projects are preferred.

Must be very smart and very motivated, must be a quick learner, proactive and curious.

Must be passionate about being part of an aggressive, venture-backed startup team that is changing the way chips are architected, designed, and programmed

Must be entrepreneurial, innovative problem solver and willing to work hard.

Must live in Silicon Valley. Strong preference for US citizenship or permanent residency (“green card”); will consider candidates with current H1-B visas who are willing to transfer promptly.