

AI Inference Software Development Manager

Flex Logix has announced its first Inference Accelerator Co-Processor, InferX X1, which is the fastest and most efficient edge inference chip. It outperforms Nvidia Xavier NX at 1/7th the die size and lower power. Production is slated for Q2/2021. Our software team is preparing our Inference Model Compiler to be ready to run deep neural network models on X1, starting with YOLOv3.

RESPONSIBILITIES

Manage about half of our growing Inference Compiler team, reporting to the Senior Director, which is responsible for our Model Compiler for our Reconfigurable Tensor Processor: a DNN Model-to-binary flow:

- Lead the team in expanding functionality of our Model Compiler, written in modern C++, for support of additional capabilities and computation optimizations
- Understand the required feature content and delegate implementation of parts to team members
- Own program management, create schedules for designs and implementations and track the team progress on a weekly basis with updates to management
- Conduct design and code reviews, promote communication and good design practices across the team

This is a hands-on management role, so in addition to understanding software you need to have some basic understanding of computer architecture, digital logic and Deep Neural Networks.

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

EXPERIENCE AND SKILL REQUIRED

8+ years of relevant industry experience and at least 3 years as a manager of a team of 3-8 software developers. AI/NN expertise is not required but desired.

Previous experience of leading the team or individual contribution into development of system-level software or embedded systems in Windows or Linux in C++ is required.

Excellent understanding of software/hardware development engineering practices, workflow, schedule dependencies.

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.