



AI Compiler Engineer

About Vsora :

VSORA is a French fabless semiconductor company delivering ultra-high-performance AI inference solutions for both data centers and edge deployments. Our proprietary architecture achieves exceptional implementation efficiency, ultra-low latency, and minimal power draw - dramatically cutting inference costs across any workload.

Fully programmable and agnostic to both algorithms and host processors, our chips serve as versatile companion platforms. A rich instruction set lets them seamlessly handle pure AI, pure DSP, or any hybrid of the two, all without burdening developers with extra complexity.

To streamline development and shorten time-to-market, VSORA embraces industry standards: our toolchain is built on LLVM and supports common frameworks like ONNX and PyTorch, minimizing integration effort and customer cost.

Based in the outskirts of Paris, France, the company was founded in 2015 by a team of highly qualified and accomplished AI/DSP experts and entrepreneurs.

Job brief :

We are seeking a highly skilled and motivated AI Compiler Engineer to lead the development of the next-generation compiler stack for Vsora's chip family. This is a hands-on leadership role where you will join a team of talented engineers, drive architectural decisions, and contribute directly to the implementation of novel compiler optimizations and fusion techniques. You'll be working at the intersection of AI and systems, helping shape the future of high-performance AI workloads on Vsora hardware.



Key Responsibilities :

- Participate to the design and development of the AI compiler stack for Vsora's chip family
- Architect and implement advanced graph fusion and optimization techniques
- Collaborate with hardware architects, ML framework teams, and runtime engineers
- Drive performance analysis and tuning across ML models and benchmarks
- Stay current with trends in AI compilers, ML frameworks, and hardware acceleration

Requirements and skills :

- Strong C++ and Python programming skills
- Deep understanding of compiler internals, IRs, and optimization pipelines
- Hands-on experience with MLIR, TVM, OpenVINO, IREE, XLA
- Familiarity with PyTorch, Tensorflow, ONNX or other ML frameworks
- Experience with performance tuning and low-level system optimization
- Excellent communication and collaboration skills

This is a plus :

- Experience with NPU, GPU, or other AI accelerator architectures
- Contributions to open-source compiler or ML projects
- Background in machine learning or numerical computing
- Familiarity with software development tools (e.g., Git, CI/CD, profilers, debuggers)

Contact :

VSORA

13/15 rue Jeanne Braconnier

Immeuble Le Pasteur
92360 Meudon-La-Fôret
France

Website : vsora.com

Email : jobs@vsora.com
