Job Title: LEAD/MANAGER AI OPTIMIZATIONS

Employment Type: Full Time

Location: Hybrid/Barcelona, Spain

Department: SW Engineering

Description: We are hiring a hands-on Software Developer Lead with experience in Generative AI to develop neural network optimizations for Machine Learning accelerators with thousands of cores. In this role, you will be responsible for defining, implementing, and validating algorithmic and compiler solutions to deliver unmatched energy efficiency at the inference stage for our HW accelerators. You will explore novel generative AI applications and adapt them to maximize the benefits of our HW architecture. As part of our dynamic SW team, you will actively participate in different components of the SW stack, such as the Compiler, the Runtime, the ML Framework, and the Firmware.

What you will do

- Define, plan, and lead software optimizations for transformer models
- Develop software-driven optimizations targeting a novel RISC-V architecture
- Implement algorithmic and compiler solutions for high-performance HW accelerators
- Propose solutions to enhance the efficiency of Machine Learning applications
- Integrate different inter-operator optimizations across the entire SW stack
- Analyze and characterize the efficiency of SW optimizations
- Interact with functional and performance simulators

Minimum Qualifications

- MS in Computer Science or a related technical field
- Strong understanding of object-oriented programming
- Knowledge of parallel programming
- Familiarity with algorithmic complexity, data structures, and graph algorithms
- Extensive experience in C++ and Python
- Working experience with Linux distributions
- Business fluent in English
- This is a hybrid position located in Barcelona

Valuable Qualifications

- Knowledge of recent C++ extensions (C++14, C++17, C++20)
- Experience with PyTorch and ONNX Runtime
- Knowledge of kernel vectorization
- Experience with basic GitLab-CI or related CI pipeline flows

- Familiarity with unittest libraries (gtest, pytest)
- Experience in ML compiler development
- Knowledge of high-performance computing
- Understanding of processor and SoC architecture
- Familiarity with AI/ML algorithms and Neural Networks (CNNs, recommending systems, transformers, LLMs, VLMs, stable diffusion, whisper)