Offer #2025-09051

Engineer F/M. MLIR based Compiler optimization framework for deep learning workloads

Contract type : Fixed-term contract

Renewable contract : Yes

Level of qualifications required : Graduate degree or equivalent

Fonction : Temporary scientific engineer

About the research centre or Inria department

The Centre Inria de l'Université de Grenoble groups together almost 600 people in 24 research teams and 9 research support departments.

Staff is present on three campuses in Grenoble, in close collaboration with other research and higher education institutions (Université Grenoble Alpes, CNRS, CEA, INRAE, …), but also with key economic players in the area.

The Centre Inria de l'Université Grenoble Alpes is active in the fields of high-performance computing, verification and embedded systems, modeling of the environment at multiple levels, and data science and artificial intelligence. The center is a top-level scientific institute with an extensive network of international collaborations in Europe and the rest of the world.

Context

CORSE is a joint research group in the LIG laboratory that regroups several expertise that stand at the interface between software and hardware: those are domain specific application/library tuning, compiler optimization, run-time systems, and debugging/monitoring. Our domains of application include performance (both speed and energy consumption), reliability, and teaching of computer science. An important activity concerns the optimization of machine learning applications for specific high-performance embedded architectures.

The position is funded by the DeepGreen project. The goal of CORSE in this project is to advance research in compiler optimization, including compiler

infrastructure and scheduling heuristics, specifically for deep learning codes.

Assignment

The aim of the position is to contribute to the development of tools that help the programmer to obtain highly optimized code of deep learning applications. This work may includes various tasks:

- Infrastructure:
  - Metaprogramming: Enrich high and low-level MLIR transform dialects with optimizing loop and hyper-block transformations
  - Traditional design: Extended MLIR with backend passes
- Language: Implementation of programers-level primitives which describe the "schedule"
- Code generation: Handle specific cases such as low-level code for sparse operators and/or data-movement/packing for distributed storage systems.

The targeted architectures are CPU-like tensor computation accelerators.

Main activities

The main activities include:

- Contributions to MLIR-based compiler infrastructure
- Development of code generators (compute & data-movement)
- Technical support to PhD students

Skills

The position requires:

1. Development experience with MLIR
2. Background in compiler infrastructures for deep-learning applications
3. Expertise in compiler optimization focusing on data locality and parallelism (including data dependencies, tiling, etc.)
4. Proficiency in C, C++, and Python programming
5. Strong communication skills (teamwork) and the ability to thrive in a research environment with flexible development directives

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)

- Possibility of teleworking (90 days / year) and flexible organization of working hours
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage under conditions

Remuneration

Gross salary : from 2 692 euros before deduction of tax incomes depending on laboral experiences and degrees.

General Information

- **Theme/Domain :** Architecture, Languages and Compilation
  System & Networks (BAP E)
- **Town/city :** Grenoble
- **Inria Center :** Centre Inria de l'Université Grenoble Alpes
- **Starting date :** 2025-10-01
- **Duration of contract :** 2 years
- **Deadline to apply :** 2025-07-24

Contacts

- **Inria Team :** CORSE
- **Recruiter :**
  Rastello Fabrice / fabrice.rastello@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Applications must be submitted online via the Inria website.
Processing of applications submitted via other channels is not guaranteed.

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.