Offer #2024-07748

PhD Position F/M Experimentation with LLMs for Fortran migration

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

Level of experience : Up to 3 years

About the research centre or Inria department

The Inria University of Lille centre, created in 2008, employs 360 people including 305 scientists in 15 research teams. Recognised for its strong involvement in the socio-economic development of the Hauts-De-France region, the Inria University of Lille centre pursues a close relationship with large companies and SMEs. By promoting synergies between researchers and industrialists, Inria participates in the transfer of skills and expertise in digital technologies and provides access to the best European and international research for the benefit of innovation and companies, particularly in the region.For more than 10 years, the Inria University of Lille centre has been located at the heart of Lille's university and scientific ecosystem, as well as at the heart of Frenchtech, with a technology showroom based on Avenue de Bretagne in Lille, on the EuraTechnologies site of economic excellence dedicated to information and communication technologies (ICT)

Context

This PhD will happen in the context of the Inria LLM4Code défi. LLM4Code is an ambitious project incorporating several INRIA groups and external partners for building reliable and productive solutions based on Large Language Models.

Assignment

During the project the phd student will focus onassessing the possibility of performing a software migration with LLMs in the specific context of a given niche technology for a given organization (specific domain, specific development culture).

Context

We are engaged with an industrial partner on a code transformation project that aims to migrate a Fortran-77 + proprietary extension code base into modern Fortran code. The project uses a model driven approach where the existing code is modeled, this model is "refactored" and then regenerated in modern Fortran.

Challenges

The performance of LLMs is correlated with their training data quality. The majority of the training dataset comes from publicly available software artifacts, and often these data can be of questionable quality, riddled with vulnerabilities, biased and produce varying outputs for identical prompts.

Generic LLMs are trained from millions of “documents”. For software engineering and code generation, specialized LLMs (like HuggingFace or Llama) have been trained, but they are bound to contain less Fortran examples as less Fortran project are available in common open-source repositories (like github).

The project will need to evaluate how such imperfect LLMs can be used for migration, what are the consequences on the quality of the result and what techniques (if any) can be used to improve these results.

Outcome

The project will propose a methodology to realize code migration of a niche technology for a specific organization using LLMs.

More importantly, it will identify the key points required in such a project and the advantages and drawback of such a project as compared for example to a deterministic model based approach?

Bibliography

- Frank F. Xu, Uri Alon, Graham Neubig, and Vincent Josua Hellendoorn. 2022. “A systematic evaluation of large language models of code”. In Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming (MAPS 2022). Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3520312.3534862

- Mahmood, Hina & Jilani, Atif & Rauf, Abdul. (2023). “Code Swarm: A Code Generation Tool Based on the Automatic Derivation of Transformation Rule Set”. International Journal of Software Engineering & Applications. 14. 1-11.

- Gustavo Pinto, Cleidson de Souza, Joa□o Batista Neto, Alberto de Souza, Tarci□sio Gotto, and Edward Monteiro, “Lessons from Building CodeBuddy: A Contextualized AI Coding Assistant”, arXiv e-prints, 2023. doi:10.48550/arXiv.2311.18450.

Main activities

Responsibilities:

- Analysis and reverse engineering of existing codebases (leveraged by Software Heritage archive)
- Applying LLM for analysis of existing code, tests and migration results
- Contributing to summarization and dissemination of results, writing scientific articles.

Skills

- Good foundation in Machine Learning and Software Engineering.
- Proficiency in OOP is required (knowing of Pharo programming language is a plus)
- Excellent problem-solving abilities and a strong interest in research.
- Ability to work independently and collaboratively in a dynamic team.
- Good communication skills (English required, French is a plus)

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 and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

2100€ gross per month for the 1st and 2nd years

2190€ gross per month for the 3rd year

General Information

- **Theme/Domain :** Architecture, Languages and Compilation
  Software engineering (BAP E)
- **Town/city :** Villeneuve d'Ascq
- **Inria Center :** Centre Inria de l'Université de Lille
- **Starting date :** 2024-10-01
- **Duration of contract :** 3 years
- **Deadline to apply :** 2024-06-30

Contacts

- **Inria Team :** EVREF
- **PhD Supervisor :**
  Safina Larisa / larisa.safina@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

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.