Offer #2023-07003

PhD Position F/M Code completion for large projects and small languages

**Contract type:** Fixed-term contract

**Level of qualifications required:** Graduate degree or equivalent

**Function:** PhD Position

### 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. It will be a joint PhD between the EVREF team and the team of R. Robbes from Labri at Bordeaux.

### Main activities

LLMs have shown very impressive results on code completion and generation, despite treating source code as text only. However, their abilities are much more advanced for languages in which data is abundant (e.g. Python, Javascript).

Some programming languages do not have that luxury, one of these being the Pharo programming language. Data is also much more abundant for popular APIs than it is for a particular project under development. This task will explore how LLMs can be leveraged for code completion, code generation, and in-IDE type inference using Pharo as an example of a language with scarce training data. We will evaluate both existing LLMs, and smaller solutions tailored to the Pharo use case. Importantly, the work will be deployed and empirically evaluated with Pharo’s user community. Since the solution will be deployed with actual users, the runtime performance of the approach will be a critical concern, potentially favoring smaller-scale and local LLMs.

We will first implement techniques that use LLMs for ranking existing entities rather than generating them [Svyatkovskiy et al., 2021] for code completion and type inference (potentially forming a feedback loop). These techniques are fast and can take into account existing identifiers in the context (since they rank them), but are limited to code completion, rather than code generation.

In a second step, we will experiment with larger LLMs in the context of Pharo. There are promising approaches to integrate project-level knowledge [Shrivastava et al., 2023b, Zhang et al., 2023]. LLMs augmented with project-level knowledge can reach the performance of much larger unaugmented models [Agrawal et al., 2023, Shrivastava et al., 2023a]. Other options include adding structural information to the LLM, such as relationships between software entities (e.g., callers/callees, data flow), or information about the runtime for type inference or to refine structural relationships (e.g., actual types of variables). Structural information and runtime information are very easily accessible in Pharo, and can be integrated for instance via adapters [Hu et al., 2021], via modifications to the attention mechanism [Hellendoorn et al., 2019], via enhanced decoding mechanisms [Agrawal et al., 2023, Geng et al., 2023] or via additional pre-training of an LLM with new training objectives [Guo et al., 2020].

In parallel, we will evaluate the practical impact of the approaches by integrating them in the Pharo IDE (with an eye towards practical performance for a code completion engine), deploying them to Pharo users, and study their impact via the analysis of telemetry data, as well as observation, interviews, or surveys [Kubelka et al., 2018].

### References
Skills

Technical
- Advanced Object-Oriented Design
- Machine learning basics
- Pharo ([http://www.pharo.org](http://www.pharo.org)) is a big plus
- Technical writing
- Synthesis and presentation

Languages:
- English
- Francais is a plus

Huma:
- Humour
- Team work
- Good communication

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
- 2082€ gross per month for the 1st and 2nd years
- 2190€ gross per month for the 3rd year

General Information
- Theme/Domain: Distributed programming and Software engineering
- Software engineering (BAP E)
- Town/city: Villeneuve d'Ascq
Inria Center: Centre Inria de l'Université de Lille
Starting date: 2024-09-01
Duration of contract: 3 years
Deadline to apply: 2024-01-31

Contacts

- Inria Team: EVREF
- PhD Supervisor: Ducasse Stephane / Stephane.Ducasse@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

CV + cover letter

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.