Offer #2024-07978

PhD Position F/M LLM-based Development of Sustainable Software Service

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

Context

Spirals is an Inria research team is the domain of distributed systems and software engineering. Spirals aims at introducing more automation in the adaptation mechanisms of software systems, in particular, transitioning from adaptive systems to self-adaptive systems. Spirals creates the future techniques for building self-healing and self-optimizing software systems.

We are offering a 3-year PhD position in the field of towards complex, self-adaptive software systems

Assignment

With the emergence of Large Language Models (LLMs), code recommenders embedded in Integrated Development Environments (IDEs) have evolved into advanced code assistants that leverage a corpus of popular code snippets, like Open Source Software (OSS) repositories, to guide the development of more and more online cloud services. These code assistants are now widely adopted by development teams and software companies thanks to the success of popular services and plugins, like GitHub Copilot. While the underlying generative models have demonstrated human-competitive capabilities to produce energy-efficient algorithms for simple problems, they fail to produce acceptable solutions for more advanced challenges, likely less prominent in training datasets [2]. This lack of diversity and expertise in code assistants may, therefore, fail to guide developers toward delivering more sustainable software services at scale. Instead, it could reinforce popular beliefs and bias [5], hence contributing to rebound effects by increasing the delivery of resource-intensive services. This is all the more damageable as the research community is investing intense efforts into assessing the energy efficiency of ICT, by studying the various levers that can influence the environmental footprint of a software service in production [1]. Therefore, influencing the generation process of code assistants in an explainable manner emerges as a critical challenge for the research community in software engineering [5]. In the context of this PhD thesis, we aim to study and promote a virtuous adoption of LLMs to control the production of sustainable software services. In particular, we intend to explore the multiple facets of LLM integrations in IDEs to assist software development teams in the adoption of environment-friendly decisions. We believe that the scope of applications for such a PhD thesis is broad, ranging from requirements elicitation, architectural design, framework adoption, code and test generation, to infrastructure configuration. We, thus, believe that such a broad scope calls for the design of sustainable LLM workflows to address the intrinsic complexity of modern software services and to reason across all the layers required to develop and operate these services.

Main activities

We can structure the activities to be addressed as part of this PhD thesis as follows:

1. Study the integration of domain-specific expertise in code assistants. This challenge intends to leverage and extend appropriate state-of-the-art approaches, such as Retrieval-Augmented Generation (RAG), or knowledge distillation, to derive expert models that favor assessed solutions in their recommendations. To do so, we intend to study alternative knowledge representations to identify the most actionable insights to be adopted in this process.

2. Study the integration of agent peers in code assistants. This challenge intends to leverage execution metrics monitored during test or production phases by PowerAPI [4] to learn about energy-efficient constructions [3]. By analyzing source code and execution trace diffs, we aim to extract code snippets that reduce the energy consumption of software and use this runtime knowledge to guide the generation of new code, or the refactoring of legacy systems.

3. Study the configuration of sustainable services with code assistants. This challenge intends to study how LLMs can be leveraged to guide the design, development, and deployment of configurable and sustainable software services.

References

[1] Bonvoisin, A., Quinton, C., and Rouvoy, R. Understanding the Performance-Energy Tradeoffs of Object-Relational Mapping Frameworks. In 31th IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER) (Mar. 2024).

[2] Coignion, T., Quinton, C., and Rouvoy, R. A Performance Study of LLM-Generated Code on Leetcode. In 28th International Conference on Evaluation and Assessment in Software Engineering (EASE) (June 2024).

[3] Danglot, B., Falleri, J.-R., and Rouvoy, R. Can We Spot Energy Regressions using Developers Tests? Empirical Software Engineering (2023).

[4] Fieni, G., Acero, D. R., Rust, P., and Rouvoy, R. PowerAPI: A Python framework for building software-defined power meters. Journal of Open Source Software 9, 98 (June 2024), 6670.

[5] Sallou, J., Durieux, T., and Panichella, A. Breaking the Silence: the Threats of Using LLMs in Software Engineering. In International Conference on Software Engineering (ICSE) - New Ideas and Emerging Results (NIER) (2024)

Skills

- Technical skills: Master's degree in IT
- Languages: French, English
- Interpersonal skills: teamwork, autonomy, taking initiative

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

General Information

- **Theme/Domain :** Distributed Systems and middleware
  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 :** 2024-08-18

Contacts

- **Inria Team :** SPIRALS
- **PhD Supervisor :**
  Quinton Clément / clement.quinton@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.