



Offer #2025-08703

Doctorant F/H LLM4Code : Coévolution continue du code pour les langages et bibliothèques grand public (LLM4Code : Continuous code co-evolution for mainstream languages and libraries)

The offer description below is in French

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

Le centre Inria de l'Université de Rennes est l'un des neuf centres d'Inria et compte plus d'une trentaine d'équipes de recherche. Le centre Inria est un acteur majeur et reconnu dans le domaine des sciences numériques. Il est au cœur d'un riche écosystème de R&D et d'innovation : PME fortement innovantes, grands groupes industriels, pôles de compétitivité, acteurs de la recherche et de l'enseignement supérieur, laboratoires d'excellence, institut de recherche technologique.

Context

La thèse s'inscrit dans le cadre du projet LLM4Code.

Assignment

La mission de cette thèse s'articule principalement autour de la réalisation d'une recherche d'excellence, que l'équipe DiverSE s'efforce de mener.

Un état de l'art fera partie des premières activités afin de mieux préparer le terrain à l'implémentation de solutions et de prototypes, ainsi qu'à la réalisation d'expériences empiriques pour une évaluation rigoureuse des contributions.

Main activities

The goal of co-evolution [Khelladi et al., 2020, Le Dilavrec et al., 2021] is to support the evolution over time of various artefacts (application code, configuration files, dependencies files, test suites, etc.). For instance, a software application needs to co-evolve due to the version upgrade of a given library or data schema. Developers must thus edit various parts of the projects while continuously ensuring that the application is still running well (e.g., through test suite execution).

LLMs can assist developers with specific related tasks integral to software co-evolution, such as code comprehension, fixes recommendation, refactoring, test evolution and augmentation, and API updates. One issue is to determine the balance between context-aware LLMs versus generic ones. For instance, GitHub's Copilot offers context-aware code suggestions, but not specifically for the software project to co-evolve. Hence, an approach is to leverage the contextual information of a software project (through analyzing data extracted from codebases, issues, programming styles, and developmental history [Le Dilavrec et al., 2023]) that can yield more accurate and relevant code suggestions than relying solely on an off-the-shelf LLM.

To address the challenges of updating the knowledge of LLMs trained on different versions of libraries, our approach is twofold. First, we aim to synthesize specific and actionable knowledge, based on a comparative analysis ("diff") between different library versions. This synthesis aims to create concise and precise information that facilitates the

LLMs' knowledge update without overloading them with voluminous data. The inadequacy of sources like StackOverflow lies in their inability to provide complete context and detailed comparison between specific versions, which is crucial for an effective knowledge update. Second, we plan to combine various information sources, such as migration examples, documentation, mailing lists, and project histories, to gain a comprehensive perspective. This multidimensional approach helps overcome the limitations of raw documentation, which often fails to explicitly compare different versions and may lack precision in code migration recommendations. By providing specific information and actionable instructions, our method aims to ease the synthesis of code adapted to the latest library versions. In our approach, Software Heritage serves as a vast repository of software development history. By mining Software Heritage, we can extract historical data, track evolutionary patterns of software libraries, and understand the context of changes over time. As part of co-evolution, we pursue related goals, like augmenting test suites or leveraging project contextual information. We plan to adopt a similar approach by synthesizing targeted "diff" knowledge and exploiting the benefits of different information sources.

This strategy is related to the concept of RAG, where the integration of external knowledge is supposed to enhance the model's generation capabilities. The specific challenge is to synthesize the precise and right amount of information as part of the RAG to then effectively co-evolve code with LLMs. An open question is how LLMs manage to reconcile potential inconsistencies between the knowledge acquired during pre-training and the newly synthesized knowledge through our approach [Luo et al., 2023, Riemer et al., 2018]. This issue of inconsistency could impact the accuracy and reliability of the LLMs, necessitating a robust mechanism to integrate updated information while maintaining coherence with their original training data. Addressing this will be crucial to ensure that the LLMs remain up-to-date and effective in handling evolving software applications. In summary, our approach is to provide relevant, precise, and tailored information to meet the specific needs of LLMs when providing code fixes or suggestions as part of

co-evolution. We plan to develop and integrate automated support for code co-evolution in mainstream, open source IDEs (e.g., VSCode).

Benefits package

- ○ ■ Restauration subventionnée
- Transports publics remboursés partiellement
- Possibilité de télétravail à hauteur de 90 jours annuels
- Prise en charge partielle du coût de la mutuelle

Remuneration

Salaire mensuel brut de 2 200 €

General Information

- **Theme/Domain** : Distributed programming and Software engineering
Software engineering (BAP E)
- **Town/city** : Rennes
- **Inria Center** : [Centre Inria de l'Université de Rennes](#)
- **Starting date** : 2025-05-01
- **Duration of contract** : 2 years, 10 months
- **Deadline to apply** : 2025-05-04

Contacts

- **Inria Team** : [DIVERSE](#)
- **PhD Supervisor** :
Khelladi Djamel / djamel-eddine.khelladi@irisa.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.

The keys to success

Vous pouvez donner là, un portrait à "gros traits" du (de la) collaborateur(trice) attendu(e) : ce que vous voyez comme nécessaire et suffisant et qui peut associer :

- goûts et appétences,
- domaine d'excellence,
- éléments de personnalité ou de caractère,
- savoir et savoir faire transversaux...

Cette rubrique permet de compléter et alléger (réduire) la liste plus formelle des compétences :

- "Se sentir à l'aise dans un environnement de dynamique scientifique, aimer apprendre et écouter sont des qualités essentielles pour réussir cette mission."
- " Passionné(e) par l'innovation, avec une expertise dans le développement Ruby on Rail et une grande capacité de conviction. Une thèse dans le domaine *** constitue un réel atout."

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

Merci de déposer en ligne CV, lettre de motivation et éventuelles recommandations

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.