



Offre n°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)

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Doctorant

A propos du centre ou de la direction fonctionnelle

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.

Contexte et atouts du poste

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

Mission confiée

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.

Principales activités

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).

Avantages

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

Rémunération

Salaire mensuel brut de 2 200 €

Informations générales

- **Thème/Domaine** : Programmation distribuée et génie logiciel
Ingénierie logicielle (BAP E)
- **Ville** : Rennes
- **Centre Inria** : [Centre Inria de l'Université de Rennes](#)
- **Date de prise de fonction souhaitée** : 2025-05-01
- **Durée de contrat** : 2 ans, 10 mois
- **Date limite pour postuler** : 2025-05-04

Contacts

- **Équipe Inria** : [DIVERSE](#)
- **Directeur de thèse** :
Khelladi Djamel / djamel-eddine.khelladi@irisa.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création

de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

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."

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

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

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.