Sujet de Stage

Titre : Secure Data Sharing Policy Recommendation with LLMs

Contexte :
Le partage de données sensibles peut être sujet à des attaques cyber de la part d'acteurs malveillants lors de collaborations multi-acteurs. Un exemple concerne l'interception de données de santé partagées entre plusieurs hôpitaux pour entraîner un modèle IA afin de revendre ces données. Un autre exemple concerne l'interception de données concurrentielles au sein d'une supply chain par un concurrent pour repositionner son prix de vente à son avantage. Une stratégie de partage des données est donc essentielle pour préserver la confidentialité des données, garantir la confiance entre les acteurs, et protéger les informations stratégiques tout en assurant le bon fonctionnement du réseau. Cependant, la sélection et la mise en œuvre des stratégies de partage sécurisé des données peuvent être complexes et prendre beaucoup de temps. Hors cette sélection doit être adaptée en fonction des exigences des utilisateurs et de l'environnement d'exécution. Il existe donc un besoin d'outils d'aide à la décision pour orienter les architectes métiers à sélectionner et à appliquer des techniques de partage de données appropriés au cas d'utilisation donné.

Objectif :
Notre laboratoire développe un outil permettant à un consortium d'acteurs souhaitant échanger des données de spécifier ces échanges grâce à une approche de model engineering. Les acteurs établissent et spécifient ensemble un processus métier. Ce processus métier sera exécuté par un smart contract pour notariser chaque échange de donnée dans la blockchain à des fins d'audit. Une étape de simulation permet d'identifier les vulnérabilités du processus afin d'améliorer sa robustesse. L'objectif de ce stage est d'ajouter à l'outil de modélisation du laboratoire un outil d'aide à la modélisation basé IA, par exemple par une approche basée LLM. Cette extension permettra de générer un processus métier à partir de documentation fournie par le consortium. Elle permettra également d'extraire la spécification des échanges données (e.g., degré de confidentialité, contraintes d'accès etc). Elle permettra enfin d'adapter le processus et la spécification générée suite à une analyse de vulnérabilités. Le stage impliquera le développement d'un prototype du système et l'évaluation de ses performances sur un ensemble de scénarios d'échange de données issus de use cases industriels.

Le/La candidat(e) retenu(e) rejoindra le Laboratoire Systèmes d'Information de Confiance, Intelligents et Auto-Organisants (LICIA) au CEA LIST.

Méthodologie :
Le/La stagiaire aura les responsabilités suivantes :

(1) **État de l'art et analyse des besoins** : Réaliser une revue exhaustive de la littérature et des solutions existantes autour de systèmes d'aide à la modélisation basés LLM dans un cadre cyber. Identifier les défis et besoins liés au cas d'usage.
(2) **Conception de l'architecture de solution** : Élaborer une architecture détaillée basée LLM pour une solution d'aide à la modélisation de systèmes de gouvernance cyber, en tenant compte des contraintes techniques et opérationnelles identifiées.
(3) **Développement du prototype** : Sur la base de l'architecture conçue, développer un prototype fonctionnel de la solution.
(4) **Tests et validation** : Effectuer une série de tests pour valider la fiabilité, la sécurité, et la performance de la solution.
(5) **Documentation et transfert de connaissances** : Rédiger une documentation complète de la solution

Compétences :
Le/La candidat(e) doit avoir les compétences suivantes :
- Etudiant(e) master 2 en informatique/ingénierie.
- Connaissance en cryptographie, sécurité informatique.
- Connaissance de l'ingénierie du logiciel (expérience préalable des modèles de développement, du cycle de vie du logiciel, ou de l'intégration continue est un atout).
- Connaissance des principes avancés de la conception du logiciel (expérience préalable des langages fortement typés, du polymorphisme, de la programmation générique, des templates, ou des design patterns est un atout).
- Expérience en IA / IA générative

Domaine de spécialité requis : Informatique

Autres domaines de spécialités, mots clés : IA générative, sécurité, systèmes distribués, conception du logiciel, ingénierie du logiciel

Moyens mis en œuvre (expériences, méthodes d'analyses, autres...) : recherche, programmation

Niveau souhaité : Bac + 5 - Master 2

Durée : 6 mois

Niveau d'habilitation défense (AS au minimum): AS

Formation souhaitée : Ingénieur/Master

Possibilité de poursuit en thèse : Oui

Lieu du stage : CEA, Centre de Saclay Nano-Innov, 91191 Gif sur Yvette

Contacts :
Tiphaine HENRY tiphaine.henry@cea.fr