Offre de stage M2 / PFE

Interprétation et raisonnement explicable sur les graphes de connaissance pour l'identification de signaux faibles

---

Contexte

Les graphes de connaissance sont aujourd'hui utilisés pour structurer et raisonner sur des masses de données complexes (santé, veille stratégique, cybersécurité, finance, etc.). Cependant, l'usage des modèles neuronaux ou d'embeddings pour l'inférence sur ces graphes pose un problème majeur d'interprétabilité : les résultats sont difficiles à expliquer et à exploiter par un humain, en particulier lorsqu'il s'agit de détecter des signaux faibles, c'est-à-dire des indices discrets, dispersés ou peu visibles mais annonciateurs d'événements futurs (ex. cyberattaques, crises, anomalies, ...).

Objectifs du stage

- Étudier et comparer différentes approches de raisonnement sur graphes de connaissance (embedding, multi-hop, causal, temporel, commonsense) en termes d'interprétabilité et d'exploitabilité [Liu et al., 2025, Wang et al., 2024, Liu et al., 2024].
- Définir des métriques adaptées à la détection de signaux faibles dans un KG (ex. rareté des relations, co-occurrence atypique, motifs causaux émergents).
- Explorer des mécanismes explicables (règles, sous-graphes, chemins multi-hop interprétables, visualisations) permettant à un analyste humain de comprendre pourquoi un signal est identifié comme faible mais significatif.
- Proposer et évaluer un prototype méthodologique sur un cas d'usage concret (ex. veille informationnelle, données d'événements temporels).
- Mettre en situation réelle les approches étudiées sur des données relles fournies par le partnaire industriel du projet.

Méthodologie envisagée

La démarche proposée s'articulera en quatre volets complémentaires. Tout d'abord, une revue de littérature permettra d'analyser les principales approches de raisonnement sur graphes de connaissances, en mettant l'accent sur leurs forces et limites en termes d'interprétabilité. Dans un second temps, une modélisation des signaux faibles sera élaborée, afin de définir des critères adaptés (relations rares, motifs causaux, événements temporels atypiques) et de proposer une typologie exploitable. La troisième étape consistera à développer un prototype expérimental comparant des méthodes de raisonnement "boîte noire" (par ex. embeddings neuronaux) avec des approches explicables (raisonnement par chemins, règles ou chaînes causales). Enfin, une phase d'évaluation combinera métriques quantitatives et critères qualitatifs d'explicabilité (fidélité, plausibilité, compréhension humaine), appuyés sur des visualisations de sous-graphes ou chemins interprétables.

Livrables attendus

- Rapport de recherche (revue de littérature + méthodologie).
- Prototype expérimental (notebook ou code Python).
- Évaluation sur un dataset réel.

Candidatures

Des compétences sont attendues en programmation et en science des données (Machine Learning et Deep Learning), en représentation des connaissances. Des connaissances en IA explicable (XAI) seront appréciées.

Profil recherché : Master 2 Informatique ou 5ème année Ingénieur.

Lieu du stage Laboratoire LIRIS, INSA Lyon, Campus La Doua, Villeurbanne, avec des déplacements ponctuels chez Auvalie Innovations à Lyon 9ème.

Période de stage : 5 à 6 mois entre février/mars et Août/Septembre 2026.

Candidature : Envoyer un mail présentant votre parcours et vos motivations, votre CV et vos derniers relevés de notes à : ludovic.moncla@insa-lyon.fr, remy.cazabet@liris.cnrs.fr et khalid.benabdeslem@univ-lyon1.fr

Références

[Liu et al., 2025] Liu, L., Wang, Z., and Tong, H. (2025). Neural-symbolic reasoning over knowledge graphs : A survey from a query perspective. *SIGKDD Explor.*, 27(1) :124–136.

[Liu et al., 2024] Liu, X., Mao, T., Shi, Y., and Ren, Y. (2024). Overview of knowledge reasoning for knowledge graph. *Neurocomputing*, 585 :127571.

[Wang et al., 2024] Wang, J., Sun, K., Luo, L., Wei, W., Hu, Y., Liew, A. W., Pan, S., and Yin, B. (2024). Large language models-guided dynamic adaptation for temporal knowledge graph reasoning. In *Advances in Neural Information Processing Systems 38 : Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*.