



## OFFRE DE STAGE

**Mission : Graph RAG pour un service d'interrogation de base documentaire métier intégrant les grands modèles de langues (LLMs).**

**Durée : 6 mois - Début du stage souhaité : mars 2025 - Lieu : EDF R&D Lab Saclay (91120)**

### Contexte et objectifs

La R&D d'EDF (2000 chercheurs) a pour missions principales de contribuer à l'amélioration de la performance des unités opérationnelles du groupe EDF, d'identifier et de préparer les relais de croissance à moyen et long terme. Dans ce cadre, le département Services, Economie, Outils Innovants et IA (SEQUOIA) est un département pluridisciplinaire (sciences de l'ingénieur, sciences humaines et sociales) qui fournit un appui à l'élaboration et au portage des offres, des services et des outils de relation client aux directions opérationnelles du groupe EDF.

Au sein de ce département, ce stage sera rattaché au groupe « Statistiques et Outils d'Aide à la Décision » (SOAD) : cette équipe compte une vingtaine d'ingénieurs chercheurs spécialisés en IA et data science avec des compétences fortes autour du machine learning et du deep learning, du web sémantique, de l'IA symbolique et de l'IA générative (texte, voix, image, multimodalité...), en particulier du NLP (LLM, RAG, data mining.).

Les Grands Modèles de Langue (Large Language Models ou LLMs) ont permis des avancées considérables dans le traitement et la génération du langage naturel. Malgré leur grand potentiel, les LLMs font face à des limites freinant leur adoption, telles que les hallucinations (réponses plausibles mais incorrectes ou infondées), ce qui compromet grandement la fiabilité des applications qui les intègrent. La **génération augmentée par récupération (Retrieval Augmented Generation ou RAG)** est une technique qui vise à améliorer la précision, la cohérence et la fiabilité des LLMs en intégrant des connaissances externes lors de la phase d'inférence. Cela permet d'inculquer des connaissances liées à un domaine spécifique à des modèles de langue généralistes par le biais de techniques de Prompt Engineering, et donc de spécialiser un modèle sans avoir à le Fine-Tuner dans ce domaine.

Cependant, les sources des connaissances utilisées sont souvent des documents non structurés avec peu de descriptions contextuelles ou de relations exprimées, ce qui représente une perte d'informations impactant la pertinence de la phase d'identification de connaissances.

Les **graphes de connaissances (Knowledge Graph ou KG)** sont des représentations structurées des entités d'un domaine ainsi que de leurs relations sémantiques. Récemment, de multiples travaux ont exploré l'usage des graphes de connaissances pour le RAG (Graph RAG) avec des implémentations présentant des améliorations significatives des performances des LLMs.

Au sein du groupe SOAD, nous travaillons sur l'exploitation des IA génératives pour l'interrogation de bases documentaires métiers. Un premier système de RAG sur une base de données interne a été développé cette année. Ce stage s'inscrit donc dans une démarche d'amélioration des performances de ce système. Il s'attachera particulièrement à évaluer l'impact du Graph RAG sur les performances du modèle comparé aux approches RAG traditionnelles déjà implémentées.

### Étapes du stage :

- Etat de l'art sur les techniques Graph RAG
- A partir de travaux initiés, contribuer à la création d'une base de graphe à partir de base de documents textuels, décrivant les connaissances métier
- Implémenter le RAG sur la base de graphe

- Evaluation des performances du LLM

### Profil recherché :

- Etudiant(e) en Master 2 ou équivalent école d'ingénieur, spécialité data science, IA ou équivalent
- Compétences solides en programmation, en particulier en Python.
- Connaissance des techniques de traitement du langage naturel (NLP) et des modèles d'IA générative (par exemple, GPT-4).
- Compétences en construction et manipulation de graphes de connaissances (RDF, Neo4j, graphDB, ontologie...) fortement souhaitées.
- Excellent niveau français oral et écrit
- Curiosité scientifique et forte motivation pour l'innovation

### Références :

- Xu, Z., Cruz, M. J., Guevara, M., Wang, T., Deshpande, M., Wang, X., & Li, Z. (2024). **Retrieval-Augmented Generation with Knowledge Graphs for Customer Service Question Answering**. *arXiv preprint arXiv:2404.17723*.
- <https://microsoft.github.io/graphrag/>
- Guan, X., Liu, Y., Lin, H., Lu, Y., He, B., Han, X., & Sun, L. (2024, March). **Mitigating large language model hallucinations via autonomous knowledge graph-based retrofitting**. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 16, pp. 18126-18134).
- Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). **Unifying large language models and knowledge graphs: A roadmap**. *IEEE Transactions on Knowledge and Data Engineering*.

## Informations pratiques

**Unité d'accueil :** Groupe SOAD (Statistique et Outils d'Aide à la Décision), département SEQUOIA d'EDF Lab Paris-Saclay, 7 boulevard Gaspard Monge, 91120 Palaiseau.

Le stage sera encadré par des ingénieurs-chercheurs Data Scientist du département SEQUOIA.

**Transmettre par mail un CV et une lettre de motivation à:**

Fatma-Zohra Hannou ([fatma-zohra.hannou@edf.fr](mailto:fatma-zohra.hannou@edf.fr)) & Leila Hassani ([leila.hassani@edf.fr](mailto:leila.hassani@edf.fr))