



DataLab Groupe Crédit Agricole

Stages 2025



Présentation du service

Crédit Agricole, Agir chaque jour dans l'intérêt de nos clients et de la société

Au sein du **Pôle Technologies, Digital et Paiements (ITD)**, la **Direction Data Groupe** a pour ambition de maximiser la contribution de **la Data et de l'Intelligence Artificielle** au fonctionnement du Crédit Agricole. Elle s'appuie pour cela sur la fonction de **Chief Data Officer Groupe**, le **Data Management Office Groupe** et le [DataLab Groupe](#), pôle de référence en **conception interne de solutions Data & IA** innovantes et industrielles en partenariat avec les entités du Groupe.

Le **DataLab Groupe** dispose de toutes les **compétences Data** coopérant au sein de **Squads pluridisciplinaires** selon une **méthode interne d'inspiration Agile**:

- **Data & AI Engineering** visant à préparer les données, définir les architectures, infrastructures et « packager » les solutions qui y seront déployées pour intégration dans le SI,
- **Data Science Analytique et Sémantique** pour concevoir des algorithmes d'Intelligence Artificielle basés sur l'open source exploitant respectivement des données structurées (tabulaires) et des données non structurée (texte, image, voix, vidéos) afin de répondre aux besoins exprimés par les métiers des entités du Groupe,
- **Gestion de projets** qui avec l'ensemble des partenaires et équipes techniques du DataLab Groupe, permet d'identifier et étudier les opportunités, cadrer les projets et en coordonner la réalisation.



Déroulement des stages

Les stages se dérouleront sous l'**encadrement d'experts IA**, au sein d'une **squad pluridisciplinaire** ayant comme **réfèrent fonctionnel un chef de projet**, dans l'objectif de livrer des fonctionnalités intégrables dans des solutions en production, selon la **méthode Projet du DataLab Groupe** qui fait l'objet d'une **certification IA de confiance et d'une labélisation RSE**.

Les étapes clés du stage sont les suivantes :

1. Veille bibliographique sur la problématique;
2. Sélection et implémentation des approches les plus adaptées à la problématique ;
3. Réalisation d'une étude comparative sur des données internes et externes;
4. Intégration des développements dans les produits du DataLab Groupe ;
5. Publication scientifique si les travaux aboutissent à de nouvelles approches plus performantes que l'état de l'art.

Le stagiaire aura accès à des environnements d'engineering industriel **puissants et à l'état de l'art**, en local et dans le cloud.

Les modèles seront évalués sur des **données internes** ainsi que des données externes (open-data), et seront intégrés dans **les produits et services IA en production**.

Des **interactions** avec l'ensemble des équipes **techniques** et des **experts métier** du Groupe auront lieu.

Liste des stages 2025

- #1 Assistant Data Scientist – Amélioration de systèmes R.A.G**
- #2 Assistant Data Engineer – Industrialisation des optimisations d'un asset de R.A.G**
- #3 Assistant Data & AI Engineer – Numérique Responsable**
- #4 Assistant Data Analyst – Risques Climatiques**
- #5 Assistant Data Scientist – Moteur de Recommandation**
- #6 Assistant Cloud/DevOps Engineer - DevOps et Ingénierie Cloud**
- #7 Assistant Chef de projet Transformation Digitale & Data (2 postes)**

#1

Amélioration de systèmes R.A.G

Assistant Data Scientist

#LLM #NLP #deep learning #RAG

Contexte du stage :

Les systèmes de génération augmentés par la recherche dans des documents (RAG) [1] jouent un rôle significatif dans le domaine bancaire. Ils permettent, entre autres, d'améliorer le support aux collaborateurs et clients au travers de **réponses composées à partir des connaissances de l'entreprise, précises, fiables et sourcées**.

Ces systèmes permettent de pallier les principales limites des LLM :

- **Hallucinations** : des informations erronées/inventées ou invérifiables par le seul contexte fourni au LLM ;
- **Informations obsolètes** : la connaissance des LLM est figée à la date de dernier apprentissage ;
- **Données d'entraînement peu fiables** : le LLM peut générer une réponse à partir de sources ne faisant pas autorité ;
- **Peu de connaissance sur des domaines très spécifiques**.

La **plateforme CA Generative Search (CAGS)** est un produit développé en interne au DataLab Groupe permettant de faire de la recherche sémantique sur différents corpus documentaires. Elle est en constante évolution afin d'améliorer les résultats. Parmi les évolutions souhaitées, on retrouve :

- **Amélioration de l'étape de recherche** en optimisant les algorithmes et techniques de filtrage et classement des résultats. Cela pourrait impliquer des techniques d'apprentissage automatique ou d'autres méthodes d'IA. [2]
- **RAG multimodal** qui utilise plusieurs types de données (ex. texte, image, son) pour améliorer la pertinence et la précision de la recherche. [3][4]
- **Finetuning de LLM** pour améliorer les performances pour des cas d'usage spécifiques à un métier ou une tâche. [5]

Objectifs du stage :

Dans ce contexte, l'objectif du stage est d'améliorer la **plateforme CAGS** en participant au **développement** de nouvelles fonctionnalités ou à l'améliorer de celles existantes.

Dans le cadre de ce stage, vous serez amené à :

- Identifier les approches applicables ;

- Benchmarker ces approches pour déterminer leur pertinence et leur efficacité ;
- Travailler en étroite collaboration avec notre équipe d'experts pour implémenter les approches choisies et participer activement à leur déploiement pour répondre aux cas d'usages métiers.

Ressources bibliographiques :

[1] Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in Neural Information Processing Systems* 33 (2020): 9459-9474.

[2] Khattab, Omar, and Matei Zaharia. "Colbert: Efficient and effective passage search via contextualized late interaction over bert." *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*. 2020.

[3] Jinhao Jiang, et al. "StructGPT: A General Framework for Large Language Model to Reason over Structured Data". *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9237–9251, Singapore. 2023.

[4] Bruch, Sebastian, Siyu Gai, and Amir Ingber. "An analysis of fusion functions for hybrid retrieval." *ACM Transactions on Information Systems* 42.1 (2023): 1-35.

[5] Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models." *arXiv preprint arXiv:2106.09685* (2021).

Pour postuler :

[STAGE - Assistant\(e\) Data Scientist – Amélioration de systèmes R.A.G H/F | Stage | Montrouge | France | Crédit Agricole Carrières \(groupecreditagricole.jobs\)](#)



#2

Industrialisation des optimisations d'un asset de recherche sémantique

Assistant Data Engineer

#LLM #NLP #deep learning #RAG #OPTIMISATION

Contexte du stage :

Les systèmes de génération augmentés par la recherche dans des documents (RAG) [1] jouent un rôle significatif dans le domaine bancaire. Ils permettent, entre autres, d'améliorer le support aux collaborateurs et clients au travers de **réponses composées à partir des connaissances de l'entreprise, précises, fiables et sourcées**.

Ces systèmes permettent de pallier les principales limites des LLM :

- **Hallucinations** : des informations erronées/inventées ou invérifiables par le seul contexte fourni au LLM ;
- **Informations obsolètes** : la connaissance des LLM est figée à la date de dernier apprentissage ;
- **Données d'entraînement peu fiables** : le LLM peut générer une réponse à partir de sources ne faisant pas autorité ;
- **Peu de connaissance sur des domaines très spécifiques**.

La **plateforme CA Generative Search (CAGS)** est un produit développé en interne au DataLab Groupe permettant de faire de la recherche sémantique sur différents corpus documentaires. Elle est en constante évolution afin d'améliorer les résultats. Parmi les évolutions souhaitées, on retrouve :

- **Connecteurs aux sources de données** comme SharePoint ou les bases de données du système d'information afin d'intégrer facilement de nouvelles données à la plateforme CAGS.
- **Streaming de la réponse** à la recherche afin que les utilisateurs commencent à voir les résultats dès qu'ils sont disponibles, sans avoir à attendre que la recherche soit intégralement terminée.
- **Méthode Asynchrone** d'appels aux LLM et de connexion aux bases de données, notamment PostgreSQL avec SQLAlchemy, ainsi que des endpoints asynchrones de FastAPI.[3]
- **Système de cache** : en mémoire avec Redis.

Objectifs du stage :

Dans ce contexte, l'objectif du stage est d'améliorer la **plateforme CAGS** en participant au **développement** de nouvelles fonctionnalités ou à l'améliorer de celles existantes.

Dans le cadre de ce stage, vous serez amené à :

- Identifier les approches applicables ;
- Benchmarker ces approches pour déterminer leur pertinence et leur efficacité ;
- Travailler en étroite collaboration avec notre équipe d'experts pour implémenter les approches choisies et participer activement à leur déploiement pour répondre aux cas d'usages métiers.

Ressources bibliographiques :

[1] Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in Neural Information Processing Systems* 33 (2020): 9459-9474.

[2] Xiaoqing Zheng et al. "Searching for Best Practices in Retrieval-Augmented Generation"

[3] <https://blog.octo.com/octo-rendre-resilient-un-projet-rag>

Pour postuler :

[STAGE - Assistant\(e\) AI Engineer – Amélioration de systèmes R.A.G H/F | Stage | Montrouge | France | Crédit Agricole Carrières \(groupecreditagricole.jobs\)](#)



#3

Numérique Responsable Assistant Data & AI Engineer

#IA #IAResponsable #Algorithmes

Contexte du stage :

Le Crédit Agricole renforce son engagement en faveur du numérique responsable visant notamment à réduire l'impact environnemental de son système d'information.

Dans ce cadre, le DataLab Groupe se concentre sur la maîtrise de l'impact des solutions d'IA sur 3 axes :

- Infrastructures
- **Algorithmes**
- Données

Pour ce stage, nous accordons une attention particulière à l'axe **Algorithmes** qui comprend trois pratiques clés :

- L'optimisation des modèles d'IA,
- La capitalisation sur des modèles existants,
- L'optimisation du code.

L'outil **open source EcoCode** a été identifié comme particulièrement intéressant dans l'optimisation du code. Il vise à rendre le code informatique plus durable. Ce projet analyse le code pour identifier et améliorer les patterns dont l'impact environnemental peut être réduit. Le stagiaire aura pour missions de proposer et développer de **nouvelles règles d'éco-conception** pour le code de solutions d'IA, d'évaluer leur efficacité, et plus largement de réaliser une veille technologique sur les outils liés au numérique responsable. Parallèlement, un état de l'art et une veille sur d'autres outils liés à l'axe des algorithmes sera également réalisé.

Ce stage offre une opportunité de contribuer activement à un projet innovant et contribuant à réduire l'impact du numérique.

Objectifs du stage :

Dans l'objectif de livrer des **fonctionnalités intégrables** au sein des **solutions industrielles en production**, les stages se déroulent sous l'**encadrement d'experts IA** au sein d'une **Squad pluridisciplinaire** ayant comme **référént fonctionnel un chef de**

projet IA, et selon la **méthode Projet** du DataLab Groupe qui fait l'objet d'une **certification** IA de confiance et d'une **labélisation** RSE.

Ressources bibliographiques :

[1] <https://ecocode.io/>

Pour postuler :

[STAGE - Assistant\(e\) Data & AI Engineer – Numérique responsable H/F | Stage | Montrouge | France | Crédit Agricole Carrières \(groupecreditagricole.jobs\)](#)



#4

Risques Climatiques Assistant Data Analyst

#ChangementClimatique #Risques #AIForGood

Contexte du stage :

Dans le cadre de son **Projet sociétal**, le Crédit Agricole, acteur engagé au sein des territoires, porte une attention forte aux **impacts du changement climatique** et à l'accompagnement de ses **clients et de la société** pour, en complément des indispensables actions de réduction des émissions de gaz à effet de serre, **anticiper les adaptations** nécessaires face aux **risques climatiques** qui s'intensifient.

De récents travaux du DataLab Groupe ont permis de projeter sur des cartes interactives ces changements climatiques en présentant par exemple **l'évolution de la température** selon les scénarios du GIEC mais également de mettre en lumière des **risques sur les rendements agricoles ou encore le bâti**. Il s'agit désormais d'aller plus loin dans cette représentation des risques notamment sur **l'habitabilité des territoires exposés aux aléas climatiques**. Ce stage permettra **d'étendre les risques déjà couverts** par les travaux du DataLab Groupe, mais également **d'accompagner des entités** que ce soit pour des besoins spécifiques ou encore pour répondre à des contraintes réglementaires.

Objectifs du stage :

Ce stage combinera des missions d'analyse de données et potentiellement de modélisation sur des risques agricoles (viticulture, pousse de l'herbe) et physiques (inondations, enneigement) :

- Exploiter les données **Open Data** permettant de représenter les impacts, valider leur pertinence et leur qualité et les mettre à disposition sur **OpenDataMart** (magasin de données Open Data interne ouvert aux entités du Groupe),
- Préparer ces données (ex. calcul d'indicateurs avancés) et les exploiter pour **Visualiser** et/ou **Modéliser**.
- **(Optionnel) Modéliser des phénomènes** en couplant les indicateurs créés ainsi que de la donnée interne (scoring, estimer un risque, etc.).

Ressources bibliographiques :

[1] Portail DRIAS : <https://www.drias-climat.fr/>

[2] European Environment Agency (2024). Responding to climate change impacts on human health in Europe: focus on floods, droughts and water quality.

[3] INRAE (2018). La vigne, le vin et le changement climatique en France.

Pour postuler :

[STAGE - Assistant\(e\) Data Analyst – Risques Climatiques H/F | Stage | Montrouge | France | Crédit Agricole Carrières \(groupecreditagricole.jobs\)](#)



#5

Moteur de Recommandation

Assistant Data Scientist

#deep learning #RecommenderSystems

Contexte du stage :

Au sein du secteur bancaire, les **moteurs de recommandation** sont des outils analytiques avancés qui aident à personnaliser les propositions d'offres, produits, services, conseils, ou contenus aux clients.

Les attentes des métiers du Marketing et de la Distribution sont fortes afin d'améliorer la pertinence des propositions et par suite l'expérience client en particulier sur les sites et applications digitales.

Objectifs du stage :

Le DataLab Groupe a réalisé de premiers travaux sur les **moteurs de recommandation** qui ont donné lieu à des **tests en conditions réelles** afin d'évaluer leur efficacité et les axes d'amélioration.

De nombreux défis restent à relever tels que le **nombre limité d'interactions** par client ou encore la **multi-bancarisation** qui complique l'analyse des comportements et préférences.

Le stage proposé s'inscrit dans l'**amélioration continue** de ces travaux.

Deux axes de travail se dégagent :

1. **Améliorer la représentation des clients** en exploitant mieux les différentes sources de données disponibles. Ces sources sont généralement des données structurées mais se présentant de manière différente, par exemple des séries temporelles à pas fixe mais aussi variable (série d'évènements) ou encore des données tabulaires (1 ligne □ 1 client).
2. **Tester des approches de recommandation reconnues** comme les méthodes de factorisation matricielle, ainsi que des approches innovantes (Sequential Recommender Systems, Deep Learning), et comparer leurs performances par rapport aux approches de scoring « classiques ».

Ressources bibliographiques :

[1] Aggregating Time Series and Tabular Data in Deep Learning Model for University Students' GPA Prediction. HARJANTO PRABOWO, ALAM AHMAD HIDAYAT, TJENG WAWAN CENGGORO, REZA RAHUTOMO, KARTIKA PURWANDARI, AND BENS PARDAMEAN

[2] DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, Xiuqiang He

[3] Time interval aware self-attention for sequential recommendation. Jiacheng Li, Yujie Wang, and Julian McAuley. In Proceedings of the 13th international conference on websearch and data mining, pages 322–330, 2020.

[4] Personalized top-n sequential recommendation via convolutional sequence embedding. Jiaxi Tang and Ke Wang. In Proceedings of the eleventh ACM international conference on websearch and data mining, pages 565–573, 2018.

Pour postuler :

[STAGE - Assistant\(e\) Data Scientist – Moteur de Recommandation H/F | Stage | Montrouge | France | Crédit Agricole Carrières \(groupecreditagricole.jobs\)](#)



#6

DevOps et Ingénierie Cloud Assistant Cloud/DevOps Engineer

#LLM #Cloud #devOps

Contexte du stage :

Les solutions d'IA répondant aux besoins des métiers du Crédit Agricole sont de plus en plus déployées dans le Cloud du fait de l'avènement des Large Language Models (LLM) dont les plus performants pour des usages très variés sont proposés depuis début 2023 au sein des écosystèmes des Cloud Providers.

La réussite du déploiement et du maintien en conditions opérationnelles de ces solutions repose sur l'industrialisation des processus, notamment le cycle de vie de génération d'images Docker (images de base, image applicative, etc.) pour une gestion plus automatisée et standardisée des environnements (développement, préproduction, production).

Objectifs du stage :

L'objectif de ce stage est de créer un **pipeline automatisé pour la génération d'images standardisées**.

Ces images serviront à la gestion d'environnements Cloud, Containers ou Machines virtuelles.

Les travaux incluront notamment :

- **La mise en place de templates et frameworks** : concevoir des modèles et architectures réutilisables (Terraform, Ansible, Docker, etc.) pour gérer les ressources et images sur cloud/ Kubernetes /VM.
- **La standardisation des processus** : industrialiser la création d'images via un pipeline CI/CD, assurant une cohérence entre les environnements de développement, de préproduction et de production.
- **L'automatisation** : automatiser création, test, et déploiement des images via des outils DevOps (GitLab CI, GitOps).
- **Le suivi des bonnes pratiques DevOps** : mettre en place et documenter des pratiques comme l'intégration continue (CI), le déploiement continu (CD), la gestion de version des images, et la gestion des artefacts.

Pour postuler :

[STAGE - Assistant\(e\) Cloud/DevOps Engineer - DevOps et Ingénierie Cloud H/F | Stage | Montrouge | France | Crédit Agricole Carrières \(groupecreditagricole.jobs\)](#)



#7

Assistant Chef de projet Transformation Digitale & Data

#Gestion de projet #IA #IA GEN

Contexte du stage :

Au cours de votre stage, vous rejoindrez l'équipe « Gestion de Projets » qui avec l'ensemble de nos partenaires et les experts Techniques du DataLab Groupe, identifie et cadre les projets puis en coordonne la réalisation.

Vous interviendrez sur les missions suivantes :

- Assistance aux chefs de projets du DataLab Groupe pour la gestion des projets (conception de supports, rédaction de compte-rendu...) incluant :
 - o Étude d'opportunité et cadrage de nouveaux cas d'usage,
 - o Pilotage de la réalisation avec les parties prenantes en assurant un reporting auprès d'un référent Chef de Projet du DataLab Groupe,
 - o Suivi du déploiement et de l'utilisation des solutions par les Métiers.
- Participation à l'animation de la communauté des experts Data & IA

Objectifs du stage :

A titre informatif, voici des exemples de livrables auxquels vous serez amené à contribuer :

- Supports de restitution d'étude d'opportunité, conventions projet avec nos partenaires.
- Supports de réunion (PowerPoint...), comptes rendus, restitutions d'ateliers d'animation, supports de communication internes ou externes.
- Documentation conformément aux normes de fabrication (certification LNE, Label IA)
- Restitutions mensuelles relatives à la gestion des activités du DataLab Groupe.

Pour postuler :

[STAGE- Assistant\(e\) Chef de projet Transformation Digitale & Data H/F | Stage | Montrouge | France | Crédit Agricole Carrières \(groupecreditagricole.jobs\)](#)

