

STAGE - GÉNÉRATION DE DONNÉES ANONYMES

H/F

Direction des données du HDH

Contrat – Stage

Rémunération – 800€ / mois

Date de début – dès que possible

Localisation - Paris 15ème (télétravail partiel possible)

LE HEALTH DATA HUB

Comment **améliorer les dépistages** et faire en sorte que les patients soient pris en charge le plus tôt possible ?

Comment leur **proposer les meilleurs traitements** sur le long cours ?

Comment **appuyer les professionnels de santé** dans un contexte clinique qui se complexifie ou en cas de crise sanitaire ?

L'**Intelligence Artificielle** et les **données de santé** font partie de la réponse. Elles sont incontournables pour la **recherche et l'innovation en santé**, par exemple, pour prévenir des insuffisances cardiaques à partir de données issues d'appareils connectés, pour accélérer le dépistage du cancer du sein à partir d'analyses automatiques des examens de mammographies ou même pour réunir assez d'informations afin **d'améliorer la prise en charge** des maladies rares.

Et pour ça, la France a la chance de disposer de bases de données extrêmement riches !

Mais ces données sont souvent sous exploitées car dispersées. Grâce à des **solutions innovantes** telles que l'IA, l'objectif du Health Data Hub est justement de permettre **d'accéder de manière facilitée, unifiée, transparente et sécurisée** à un catalogue de bases de données de santé françaises.

Comment ?

Le Health Data Hub a mis en place une **plateforme technologique** qui met à disposition des porteurs de **projets d'intérêt public**, dans un environnement technologique sécurisé et à l'état de l'art, les données de santé **pseudonymisées** des français. Ces porteurs de projets vont mobiliser des sources de données très **volumineuses**, les croiser entre elles, et utiliser une puissance de calcul pour faire tourner des algorithmes de recherche complexes. Il s'agit par exemple de projets de start-up pour améliorer des logiciels d'aide au professionnel de santé, de projets permettant **d'améliorer la prise en charge des patients** en comparant l'efficacité de prise en charge, de projets portés par les administrations pour éclairer les politiques publiques.

Notre offre technologique, en constante évolution, peut être consultée [ici](#).

Les défis sont de taille pour traiter ces données de santé sensibles, volumineuses de natures et formats variables. La plateforme doit être un **levier d'innovation** dans l'écosystème de la donnée de santé français.

En résumé, avec le Hub, nous **accompagnons des porteurs de projets innovants** qui contribuent à trouver les solutions de demain pour améliorer la santé de tous les citoyens.

DESCRIPTIF DU POSTE

Direction des données :

Pour mener à bien les missions qui lui ont été confiées, le Health Data Hub a formé la direction des données dont les objectifs principaux sont de :

1. **Soutenir les projets d'intérêt public** que le HDH accompagne, aussi bien sur la compréhension des données de santé que sur leur exploitation via des experts des données de santé, des data scientists et des data engineers ;
2. **Gérer et mettre à disposition les données** qui lui sont confiées aux porteurs de projet au sein de la plateforme technologique du Health Data Hub ;
3. **Partager et mutualiser les outils et les connaissances** nécessaires à l'analyse des données de santé, dans le cadre d'une démarche open source ;
4. **Définir des stratégies novatrices** sur la gestion, l'exploitation et le partage de données de santé, permettant de réaliser la vision du HDH.

Au sein de la Direction des données, vous serez rattaché au pôle "d'accompagnement des projets" et aurez la mission d'aider au développement d'outils de génération de données permettant de fournir de nouveaux moyens de formation et de mesurer la faisabilité d'une étude sur les données de santé d'origine.

Mission du stage :

Les données de synthèse sont un outil précieux pour les projets de recherche. Ces données, similaires aux données réelles mais générées de façon aléatoire, peuvent être obtenues avant l'accès aux données réelles, généralement personnelles et dont l'accès est restreint par le RGPD, si elles sont anonymes. Elles permettent de découvrir une base de données, de construire du matériel pédagogique, de réaliser quelques études sommaires comme des études de faisabilité, ou encore d'écrire des programmes sans avoir accès aux données, permettant ainsi aux chercheurs de se former et de gagner un temps précieux.

L'objet du stage est de créer des générateurs de données de synthèse pour un certain nombre de types de données rencontrées dans les projets de recherche: images, texte, signaux biologiques, etc. sous la forme de packages Python qui seront mis à disposition de l'écosystème en open source via le [Gitlab du HDH](#).

Concrètement, il s'agira :

- d'identifier les types de données les plus stratégiques se prêtant à l'exercice ;
- d'identifier les jeux de données ou publications déjà disponibles s'il en existe ;

- de créer des mécanismes d'agrégation, ou bien d'entraîner des modèles génératifs type GAN ou autoencodeurs variationnels, ou encore de développer des générateurs *ex nihilo* selon les cas ;
- et de produire un ou des packages en open source.

PROFIL RECHERCHÉ

Compétences techniques

- Python
- Git
- Deep learning

Qualités relationnelles et personnelles

- Motivation
- Rigueur, clarté de présentation
- Sens du travail en équipe
- Force de proposition
- Curiosité intellectuelle

POURQUOI CHOISIR LE HEALTH DATA HUB ?


Vous êtes motivé(e) à rejoindre une équipe impliquée dans un projet ambitieux, qui a du sens et une finalité d'intérêt public ? Rejoignez-nous !


Notre structure récente, d'une cinquantaine de collaborateurs/trices, a besoin de talents créatifs, autonomes et proactifs pour continuer de grandir !


Ensemble, nous nous sommes engagés à :

- Accompagner les porteurs de projet visant à analyser les données de santé pour le bien commun.
- Construire et opérer une plateforme technologique pour leur offrir les meilleurs outils avec un très haut niveau de sécurité à respecter.
- Réunir et mettre en forme les données au plus grand potentiel pour la recherche et l'innovation.
- Promouvoir le partage des connaissances, des expertises et du savoir et diffuser une culture de la donnée de santé auprès de tous.

Bon à savoir:

 Rejoindre le HDH c'est surtout participer à un projet enrichissant humainement qui a du sens, avec un fort impact sociétal

 Au HDH on favorise la prise d'initiative, dans une ambiance de challenge perpétuel

 Ici la bonne humeur et l'esprit d'équipe règnent

PROCÉDURE DE RECRUTEMENT

Après avoir postulé, le recrutement se déroulera de la manière suivante :

- Un premier entretien avec un membre de l'équipe (45 min)
- Une mise en situation à réaliser chez soi et un compte rendu des échanges
- Un entretien avec le manager basé sur la mise en situation (45 min)
- Une rencontre RH

Pour postuler

Retrouvez nous [ICI](#) sur Welcome to the Jungle (CV et Lettre de motivation)