



# ENVIRONNEMENT ET CONTEXTE DE TRAVAIL

Notre établissement fait partie de l'Université PSL. Située au cœur de Paris, celle-ci fait dialoguer tous les domaines du savoir, de l'innovation et de la création. Classée parmi les 50 premières universités mondiales, elle forme au plus près de la recherche des chercheurs, artistes, ingénieurs, entrepreneurs ou dirigeants conscients de leur responsabilité sociale, individuelle et collective.

## STRUCTURE D'ACCUEIL

### Environnement de travail

Le [Collège de France](#) est un grand établissement public d'enseignement supérieur et de recherche. Institution unique en France et sans équivalent à l'étranger, le Collège de France répond à une double vocation : être à la fois le lieu de la recherche la plus audacieuse et celui de son enseignement. Voué à la recherche fondamentale, le Collège de France possède cette caractéristique singulière : il réalise puis enseigne « le savoir en train de se constituer dans tous les domaines des lettres, des sciences ou des arts ».

Situé sur différents sites de Paris (place Marcelin Berthelot, rue du Cardinal Lemoine, rue d'Ulm, Belle Gabrielle) l'établissement héberge un millier de personnes : enseignants-chercheurs, chercheurs, doctorants et post-doctorants, ingénieurs et techniciens, bibliothécaires, administratifs.

Le Collège de France est membre associé de l'Université Paris Sciences et Lettres (PSL).

### Contexte

Le Collège de France est à la fois un centre de recherche de très haut niveau et un lieu ouvert à tous ceux ayant soif de connaissance. Dernièrement, l'institution a lancé le projet : « **Agir pour Éducation - Un enjeu scientifique pour la société** » Dans le cadre de cette initiative, le professeur Stanislas Dehaene a lancé un projet pour l'application des sciences cognitives à l'éducation, appelé Excello.

**L'association à but non lucratif EXCELLO** a vocation à développer, expérimenter, et diffuser en milieu scolaire des nouveaux outils d'apprentissage fondés sur les sciences cognitives. Nous sommes passionnés par l'accélération de l'apprentissage par le biais des meilleures pratiques qui sont les plus conformes à la façon dont le cerveau de l'enfant apprend.

Notre culture est délibérément axée sur les résultats. En tant qu'organisme à but non lucratif, nous avons la liberté de développer et de tester nos produits, et nous ne les mettons sur le marché qu'une fois qu'ils ont fait leurs preuves en classe. Par ailleurs, notre travail est développé en utilisant des outils open-source pour un partage facile avec la communauté des chercheurs et des éducateurs. <https://excellolab.org>

## A propos de nous

Excello a pour mission de transformer l'apprentissage de la lecture des enfants à travers le monde grâce à Kalulu, une méthode phonétique conçue pour fonctionner dans n'importe quelle langue. Notre objectif est de permettre aux élèves, parents et enseignants de suivre les progrès et l'utilisation des applications de jeu Kalulu, du matériel pédagogique imprimable et des contenus éducatifs. Pour atteindre cet objectif, nous recherchons un Ingénieur Systèmes de Données chargé de concevoir une infrastructure sécurisée et mondiale pour la collecte de données, la gestion des utilisateurs et les tableaux de bord analytiques. <https://kalulu.excellolab.org/>

# MISSIONS

## ACTIVITÉS PRINCIPALES

### Mission

Rôle central dans la construction de l'infrastructure de données de Kalulu : mise en place d'une collecte de données sécurisée et respectueuse de la vie privée (in-app et en ligne), création de systèmes utilisateurs basés sur des comptes, gestion de flux de données, mise en place de tests A/B et visualisation des données sur un tableau de bord centralisé pour utilisateurs, parents et administrateurs. Nécessite des compétences en back-end, data engineering, DevOps et une solide compréhension des réglementations de protection de la vie privée.

### Activités principales

#### Reconnaissance vocale de l'enfant

- Développer et adapter des modèles de reconnaissance automatique de la parole (ASR) spécifiquement entraînés sur des voix d'enfants (3-12 ans), en tenant compte des particularités phonétiques, prosodiques et articulatoires de cette population.
- Mettre en œuvre des pipelines de traitement audio in-app capables de fonctionner en temps réel et hors ligne (stockage local chiffré, synchronisation différée).
- Concevoir des protocoles de collecte vocale standardisés pour une utilisation dans des contextes de test et d'évaluation des apprentissages.
- Évaluer les performances des modèles (WER, CER) sur des corpus multilingues et multi-accentés d'enfants, et contribuer à l'amélioration continue des modèles.

#### Collecte de données à grande échelle depuis l'application

- Construire un système de collecte de données in-app haute fiabilité (événements d'interaction, enregistrements vocaux, métriques de progression) fonctionnant offline-first avec synchronisation serveur.
- Concevoir des flux de données sécurisés (ETL) pour traiter, anonymiser et stocker les données utilisateurs — en particulier les données vocales — en garantissant chiffrement TLS/SSL en transit et AES-256 au repos.
- Mettre en place une infrastructure d'ingestion capable de gérer des volumes importants de données multimodales (audio, logs, métadonnées) provenant d'un grand nombre d'utilisateurs simultanés.

#### Communication avec la plateforme de visualisation

- Concevoir et implémenter des APIs (REST ou GraphQL) permettant la transmission structurée des données collectées depuis l'application vers une plateforme centralisée.
- Développer des connecteurs et pipelines de données permettant l'alimentation en temps réel ou quasi-réel de tableaux de bord destinés aux enseignants, parents et administrateurs.

### **Compétences techniques**

- Maîtrise de frameworks tels que Kaldi, ESPnet, Whisper, wav2vec 2.0 ou équivalents. Expérience avec des corpus vocaux infantiles (MyST, CMU Kids, etc.). Traitement du signal audio & ASR
- Python (FastAPI, Flask) ou Node.js. Conception d'APIs robustes pour la transmission de données structurées. Back-end & APIs
- AWS (S3, Lambda, DynamoDB ou équivalents), pipelines ETL, gestion de flux temps réel (Kafka, Kinesis ou équivalents). Infrastructure de données
- Maîtrise des mécanismes de stockage local (SQLite, Realm, fichiers chiffrés) et de synchronisation différée. Collecte offline
- Chiffrement AES-256, TLS/SSL, RGPD, COPPA, anonymisation de données vocales. Sécurité & conformité

ICI Vous pouvez retrouver l'annonce complète

## PROFIL DU CANDIDAT

## SAVOIRS ET COMPÉTENCES ATTENDUS

### Formation & expériences

- Doctorat en traitement automatique de la parole, linguistique computationnelle, apprentissage automatique ou domaine connexe.
- Expérience démontrée en reconnaissance vocale de l'enfant ou en adaptation de modèles ASR à des populations spécifiques.
- Expérience en collecte et traitement de données à grande échelle depuis des applications mobiles ou web.
- Une expérience dans le domaine EdTech ou dans des applications impliquant des données enfants est un atout fort.

### Modalités de candidature

Le dossier de candidature, constitué des documents suivants :

- Une lettre de motivation
- Un curriculum vitae précisant l'employeur et la situation statutaire

Il doit être adressé dans un délai de 4 semaines suivant la publication à la Direction des Ressources Humaines à l'adresse suivante

### NON DISCRIMINATION, OUVERTURE ET TRANSPARENCE

Notre établissement, comme l'ensemble de l'Université PSL, s'engage à soutenir et promouvoir l'égalité, la diversité et l'inclusion au sein de ses communautés. Nous encourageons les candidatures issues de profils variés, que nous veillerons à sélectionner via un processus de recrutement ouvert et transparent.

## CONTACT

p4bhf1e1mu5k@emploi.beetween.com

## AUTRES INFORMATIONS

Type de contrat / de poste : **CDD**

Durée du contrat : **12 mois**

**Référence**  
**p4bhf1e1mu**

# L'Université PSL (Paris Sciences & Lettres)

