

Post-doctorant F/H

Collecte de données de nos applications, A/B Testing & conformité, reconnaissance vocale

Catégorie	A
Modalités de recrutement	Contractuel CDD 1 an (salaire en fonction de l'expérience professionnelle)
Localisation	Collège de France Chaire du professeur Dehaene 11 place Marcelin Berthelot 75005 Paris
Projet	Agir pour Éducation – Un enjeu scientifique pour la société projet pour l'application des sciences cognitives à l'éducation, appelé Excello
Date de prise de fonction souhaitée	1 ^{er} juin 2026
Modalités de candidature	CV et Lettre de motivation job-ref-1oaa4zd96l@emploi.beetween.com
Date de publication	10 mars 2026

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/>

Mission

Vous jouerez un rôle central dans la construction de l'infrastructure de données de Kalulu. Vous serez responsable de la mise en place d'une collecte de données sécurisée et respectueuse de la vie privée dans les environnements in-app et en ligne. Votre rôle comprend la création de systèmes d'utilisateurs basés sur des comptes, la gestion de flux de données, la mise en place de fonctionnalités de test A/B et la garantie que les utilisateurs, les parents et les administrateurs peuvent visualiser les données sur un tableau de bord centralisé. Ce rôle nécessite un mélange d'ingénierie back-end, d'ingénierie des données et de compétences DevOps, ainsi qu'une solide compréhension des réglementations en matière 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-accent d'enfants, et contribuer à l'amélioration continue des modèles.
- Assurer la robustesse du système face aux conditions acoustiques variables (bruit de fond, environnements scolaires ou domestiques).

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.
- Collaborer avec les équipes produit et learning science pour définir les événements à instrumenter et les métriques clés d'engagement et de performance.

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

Compétences comportementales

- Attention aux détails : Précision dans le traitement des données sensibles des élèves et respect de la vie privée.
- Résolution de problèmes : Capacité à diagnostiquer et à résoudre les erreurs du système, en particulier lorsqu'il s'agit de tests A/B ou de données d'utilisateurs.
- Collaboration inter-équipes : Travailler en étroite collaboration avec des scientifiques de l'apprentissage, des spécialistes de la confidentialité des données et des gestionnaires de produits afin de garantir l'exactitude des données et une conception axée sur la protection de la vie privée.

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.

Pourquoi nous rejoindre ?

Une mission à relever : Participez à une mission visant à aider les enfants du monde entier à apprendre à lire grâce à une plateforme éducative mondiale et multilingue. Vous travaillerez avec des scientifiques et des passionnés d'éducation sur des projets internationaux en Europe et en Amérique du Sud.

Propriété : Vous serez responsable de la fondation d'un système clé qui alimente les comptes d'utilisateurs, le suivi des données et les tableaux de bord.

Innovation : Vous travaillerez sur une infrastructure de test A/B de pointe et construirez des outils pour mesurer et optimiser les résultats d'apprentissage.

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 : job-ref-1oaa4zd96l@emploi.beetween.com

Notre établissement, 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.