



## Stage en informatique / machine learning (4 à 6 mois)

Contexte : apprentissage de langues étrangères assisté par ordinateur

**Encadrement:** Isabelle Ferrané, Thomas Pellegrini, Julien Pinquier (IRIT, France), Lionel Fontan (Archean LABS, Montauban, France).

### Contexte

**ALAIA (Apprentissage des Langues assisté par Intelligence Artificielle)**<sup>1</sup> est un laboratoire commun entre la PME Archean Technologies (Montauban, France) et l'IRIT (Institut de Recherche en Informatique de Toulouse). Ce projet est soutenu par le programme LabCom de l'Agence Nationale de la Recherche (2019-2023). Il réunit les compétences complémentaires des partenaires académiques et industriels pour le développement d'applications d'apprentissage des langues étrangères selon une approche multidisciplinaire combinant pédagogie des langues étrangères, linguistique et science des données.

Dans le cadre du développement **d'exercices de prononciation en français**, vous aurez en charge le **développement d'un système permettant d'identifier automatiquement, dans des enregistrements effectués par des étudiants japonais apprenant le français, les segments correspondant à certains sons cibles de la langue apprise.**

### Description de la mission

La mission consistera plus précisément à développer, par des techniques de **traitement du signal** et de **machine learning**, un système prenant en entrée **deux enregistrements audio d'un même mot** :

- le mot prononcé en français par une personne francophone native (mot "bien prononcé")
- le même mot prononcé par un apprenant japonophone (mot contenant potentiellement des "erreurs" de prononciation, annotées au préalable par des experts)

Le système devra être capable de mettre en correspondance les sons (ou groupes de sons) prononcés par l'étudiant japonais avec les sons attendus.

Les performances de ce système devront être évaluées et comparées à celles obtenues avec un système de reconnaissance automatique de la parole.

Ce travail nécessitera de collaborer avec les personnes recrutées sur le projet (ingénieur, doctorants) ainsi qu'avec des ingénieurs travaillant pour le partenaire industriel du projet. En effet, le but d'ALAIA est de concevoir, développer et intégrer des composants logiciels sous forme de services web sur la plateforme d'apprentissage des langues déployée par Archean Labs.

---

<sup>1</sup> <https://www.irit.fr/SAMOVA/site/projects/current/labcom-alaia/>



### **Formation initiale**

- 2ème ou 3ème année d'école d'ingénieur / Master 1 ou 2 en informatique

### **Compétences attendues**

- **Compétences techniques :**
  - solides connaissances en programmation (Python, Shell), en traitement du signal (si possible de parole) et en apprentissage automatique (développement et évaluation)
  - connaissance des boîtes à outils d'apprentissage automatique (Sklearn, PyTorch, Tensorflow...)
- **Autres compétences:**
  - aptitude pour le travail en équipe
  - bonne capacité d'organisation et de restitution de son travail
  - bonnes pratiques de développement informatique (documentation, versionnage et archivage)

### **Conditions de travail**

- **Lieu de travail :** IRIT - 118, route de Narbonne 31062 TOULOUSE, avec éventuellement quelques réunions ponctuelles à Montauban (82000)
- **Durée de la mission :** 4 à 6 mois en fonction du niveau de formation concerné.
- **Date de début :** début mars ou avril 2022
- **Rémunération :** gratification de stage selon les tarifs en vigueur.

### **Candidature**

Les candidatures composées d'un **CV détaillé** et d'une **lettre de motivation** sont à envoyer aux adresses suivantes : [isabelle.ferrane@irit.fr](mailto:isabelle.ferrane@irit.fr) et [fontan@archean.tech](mailto:fontan@archean.tech)