

## FICHE DE POSTE : stage d'ingénieur TAL

*Cette fiche de poste reste indicative et les activités qui le composent sont susceptibles de varier en fonction de l'évolution des connaissances du métier et des nécessités de service. Tous nos postes sont ouverts, à compétences égales, aux candidatures de personnes reconnues travailleurs handicapés et autres bénéficiaires de l'obligation d'emploi.*

### CONDITIONS D'EXERCICE

SITE / LIEU	Maison SHS Annie Ernaux (site des Chênes) 33 Boulevard du Port 95000 Cergy FRANCE
SPÉCIFICITÉS	Temps plein
TYPE DE CONTRAT	Stage
DATE DE DÉBUT	01/02/2026
DURÉE	5 mois
LABORATOIRE DE RATTACHEMENT	Lexiques, Textes, Discours, Dictionnaires : Centre Jean Pruvost - EA 7518
RESPONSABLE SCIENTIFIQUE	Marine Delaborde
GRATIFICATION	Basé sur la législation en vigueur (environ 620€ par mois)

### DESCRIPTIF DU POSTE

CONTEXTE	<p>Le stage se déroulera au laboratoire <a href="#">LT2D</a> (Lexiques, Textes, Discours, Dictionnaires : Centre Jean Pruvost - EA 7518) : Les chercheurs du LT2D développent des projets de recherche originaux dans les domaines du lexique, des textes, des discours et des dictionnaires. Dans la lignée de l'équipe Métadif et de l'UMR LDI, le laboratoire s'inscrit dans la stratégie de recherche et d'innovation de CY Cergy Paris Université par l'ancrage de ses recherches dans les thématiques transversales « Transition sociétale et environnementale » et « IA et applications ».</p> <p>Le stage se déroulera dans le cadre du projet lié à la CPJ « Ressources numériques en Sciences Humaines et Sociales ». Ce projet a pour objectif de valoriser les ressources linguistiques du laboratoire ainsi que d'en créer de nouvelles.</p>
----------	---

	<p>Le stage permettra d'apporter une contribution à la création de nouvelles ressources en proposant des pistes d'enrichissement pour un outil en cours de développement : <a href="#">AliGator</a> (Dehouck et Delaborde, 2023 ; 2025), un générateur automatique de jeux de mots à base de prénoms. Cet outil a été pensé comme une alternative à la fois moins gourmande et plus facilement interprétable que les LLMs dont la capacité à lier écrit et oral reste encore largement perfectible. AliGator permet d'envisager des utilisations en didactique, notamment à travers le développement de ressources matérielles.</p>
MISSIONS ENVISAGÉES	<p>La personne recrutée pour ce stage participera à l'intégration de modèles de langue (LLMs) dans l'outil AliGator dans le but d'améliorer la cohérence sémantique des jeux de mots.</p> <p>AliGator génère des jeux de mots en alignant les transcriptions phonétiques de mots et d'expressions en français avec celles de prénoms. L'ajout de règles de composition syntaxique permet d'étendre énormément les possibilités de création de jeux de mots. Cependant la plupart de ces jeux de mots ne sont pas drôles car ils n'ont pas de sens.</p> <p>L'objectif de ce stage sera donc d'associer les capacités sémantiques des modèles de langues à l'architecture syntaxico-phonétique de AliGator pour créer de meilleurs jeux de mots.</p>
PROFIL RECHERCHÉ	<p>La personne recrutée devra être inscrite dans une formation telle qu'un Master ou une école d'ingénieur avec une spécialisation en informatique, intelligence artificielle, traitement automatique des langues, linguistique informatique ou dans un domaine connexe.</p> <p>Compétences :</p> <ul style="list-style-type: none"> <li>- maîtriser les concepts de machine learning et deep learning ;</li> <li>- avoir une bonne connaissance de Python et une expérience avec la librairie Transformers ;</li> <li>- être à l'aise avec la lecture d'articles scientifiques, notamment en anglais, et savoir effectuer une veille scientifique pour se tenir au courant des actualités en TAL ;</li> </ul> <p>Une expérience avec les sciences du langage et notamment la syntaxe et la phonétique ainsi qu'un attrait pour les jeux de mots seront fortement appréciés.</p>
RÉFÉRENCES BIBLIOGRAPHIQUES	<ul style="list-style-type: none"> <li>- Dehouck, M., &amp; Delaborde, M. (2023). Génération automatique de jeux de mots à base de prénoms. In <i>18e Conférence en Recherche d'Information et Applications, 16e Rencontres Jeunes Chercheurs en RI, 30e Conférence sur le Traitement Automatique des Langues Naturelles, 25e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues</i> (pp. 1-2). ATALA.</li> <li>- Dehouck, M., &amp; Delaborde, M. (2025, January). Rule-based Approaches to the Automatic Generation of Puns Based on Given Names in French. In <i>Proceedings of the 1st Workshop on Computational Humor (CHum)</i> (pp. 18-22).</li> </ul>

## POUR POSTULER

CONTACT	Pour postuler, ou pour toute question à propos du stage, vous pouvez envoyer un mail à : <a href="mailto:marine.delaborde@cyu.fr">marine.delaborde@cyu.fr</a> en mettant <a href="mailto:mathieu.dehouck@cnrs.fr">mathieu.dehouck@cnrs.fr</a> en copie.
DOCUMENTS À JOINDRE	<ul style="list-style-type: none"><li>- CV</li><li>- lettre de motivation</li><li>- relevé de notes de l'année en cours et/ou passée</li></ul>
DATE LIMITE	2 novembre 2025