2026–2027
Sujet de stage M1/M2
Titre: Recherche Interdisciplinaire en Intelligence Artificielle sur les Archives Autochtones d'Alaska

Contexte

Ce stage interdisciplinaire est particulier. Il est ouvert à toutes les disciplines et peut être orienté davantage vers l'informatique et les mathématiques, vers la linguistique ou encore vers la sociologie et l'anthropologie. Dans toutes les déclinaisons, le/la stagiaire devra s'insérer dans une équipe de recherche très interdisciplinaire.
L'objectif principal de ce stage est d'étudier les informations sur les variations climatiques dans l'Arctique cachées dans les récits de prêtres orthodoxes en Alaska aux XIX et XX siècles.
L'Alaska, maintenant le 49ème État des États-Unis, a été colonisée à partir du début du XIX siècle par ces missionnaires de l'Église orthodoxe russe. Ces prêtes-missionnaires, d'origine mixte, étaient chargés de plusieurs communautés autochtones d'une certaine région de l'Alaska. L'Alaska ayant une superficie environ 3.1 fois celle de la France métropolitaine avec une population totale ne dépassant pas 600 milles personnes (en 2026), cette charge confiée à chaque prêtre impliquait une forte nécessité de se déplacer constamment sur de grandes distances. Lors de ces déplacements, les prêtres notaient toutes leurs observations dans des journaux.
Ces journaux, conservés en partie à la Library of Congress et en partie sur place en Alaska, constituent les premiers points de données historiques pour les conditions climatiques et environnementales en Alaska. Les cultures autochtones sont basées sur la narration et non l'écrit; elles n'avaient donc pas gardé ce savoir historique pour une analyse moderne par écrit.
En somme, les archives de la Library of Congress représentent environ 60 000 pages manuscrites avec une variante de l'alphabet cyrillique ainsi qu'une orthographe russe ancienne. Une analyse manuelle de ces données brutes est donc hors de toute atteinte.
Dans le cadre de ce stage, l'étudiant(e) poursuivra la conception et le développement d'une chaîne de traitement et d'analyse de ces récits historiques en s'appuyant sur les outils qu'offrent l'intelligence artificielle et les sciences des données [1]. A l'issue de la chaîne, on s'attendra à des données représentées sous forme d'histogrammes, analysées par les méthodes statistiques pour leur significativité et utilisables pour une comparaison avec les observations climatiques modernes, pour fournir une base pour la recherche en climatologie.
Les journaux des prêtres étant à valeur communautaire importante pour les peuples autochtones d'Alaska à nos jours, cette recherche se fait et doit se faire dans le respect absolu des valeurs autochtones. Une coopération avec une chercheure en Anthropologie constitue une partie clef de ce stage interdisciplinaire. La recherche est faite dans l'esprit du modèle "Co-production des savoirs" ("co-production of knowledge"), accepté en Anthropologie comme un modèle de recherche éthique.
L'étudiant(e) qui candidate à ce stage n'est pas obligé(e) d'avoir des connaissances dans tous les sous-domaines auxquels touche ce stage: informatique, sciences des données, anthropologie, langues russe et autochtones. Mais il ou elle doit être prêt(e) à travailler de manière interdisciplinaire avec une ouverture d'esprit et une curiosité pour la découverte. Ce stage peut être très enrichissant mais peut être également demandeur en investissement.

Objectifs

L'objectif principal de ce stage est la conception d'une chaîne de traitement d'informations par une intelligence artificielle. Le travail pourra être organisé selon les étapes suivantes :

- Segmentation des pages brutes numérisées.
- Reconnaissance des mots et caractères manuscrits en russe et langues autochtones.
- Analyse par intelligence artificielle et Techniques des Sciences des Données.
- Visualisation des Résultats.
- Réentraînement itératif des modèles afin d'améliorer leurs performances..

La recherche en informatique et en intelligence artificielle réalisée durant ce stage doit s'insérer dans une approche interdisciplinaire en travaillant étroitement avec les anthropologues, les autochtones et les chercheurs en hydrologie et climatologie.
Les travaux du/de la stagiaire pourront donner lieu à des publications dans des conférences et revues nationales et internationales, ainsi qu'à des présentations lors des séminaires de l'équipe. La soutenance de stage aura lieu à l'institution d'origine de l'étudiant(e), avec une présence des encadrant(e)s de stage en visioconférence. Les encadrant(e)s américain(e)s délivreront un certificat de réussite de stage détaillé, ainsi que des lettres de recommandation pour la suite des études du/de la stagiaire. Les stages M2 peuvent donner lieu à des propositions de continuation en thèse, sous condition d'adéquation des compétences, des financements disponibles et des règles d'admission de l'école doctorale locale.

Etablissement d'accueil

L'Université du Texas à El Paso (UTEP) est une université publique américaine classée **R1** (à activité de recherche de haute catégorie) par la classification Carnegie, qui distingue les établissements ayant une très forte activité de recherche. Située au Texas, à la frontière entre les États-Unis et le Mexique, elle accueille plus de 26000 étudiants et propose 169 formations de licence, de master et de doctorat. Elle se distingue par son engagement en faveur de l'égalité des chances et de l'accès à un enseignement supérieur de qualité. L'UTEP figure parmi les 5% d'universités américaines les mieux classées en matière de recherche. Son département d'informatique, rattaché au *Miguel A. Loya College of Engineering*, propose des formations en informatique du niveau licence au doctorat, et collabore avec des partenaires académiques et industriels. Ses principaux axes de recherche couvrent le génie logiciel, la cybersécurité, l'intelligence artificielle et les fondements théoriques de l'informatique.

Visa et mobilité

La réalisation du stage à l'Université du Texas à El Paso (UTEP) nécessite l'obtention d'un visa J-1. L'établissement pourra accompagner le/la candidat(e) dans ses démarches administratives, mais les frais liés à la demande de visa et aux formalités associées resteront à sa charge. L'accueil en stage est conditionné à l'obtention du visa requis.

Durée, rémunération et frais de scolarité

Le stage, d'une durée de quatre à six mois, n'est pas rémunéré. Aucun frais de scolarité (tuition fees) ne sera demandé par l'UTEP au stagiaire dans le cadre de ce stage, et une attestation de stage sera délivrée à l'issue de ce dernier (cf. ci-dessus). Les encadrant(e)s aideront le/la stagiaire pour toute démarche pour obtenir une bourse de mobilité internationale.

Encadrement

Le (la) stagiaire sera encadré(e) par :

- Christoph Lauter – Maître de conférences en informatique (Associate Professor), HDR, à l'Université de l'UTEP (`https://www.christoph-lauter.org/`)
- Martine Ceberio – Professeure en Informatique à l'UTEP (`https://expertise.utep.edu/profiles/mceberio`)
- Olga Lauter – Maîtresse de conférences d'enseignement (Assistant Professor of Instruction) en Anthropologie à l'UTEP (`https://www.olga-lauter.org/`)
- Hanane Benmaghnia – Post-Doctorante en validation numérique au LIP6, Sorbonne Université – CNRS, future Chargée de recherche à l'UTEP, (`https://hanane-benmaghnia.github.io/`)

Profil recherché

Le stage s'adresse à des étudiant(e)s (M1 ou M2) en informatique, mathématiques appliquées ou dans un domaine connexe en premier lieu, mais est ouvert aussi aux étudiant(e)s d'autres disciplines s'ils/elles peuvent démontrer une forte capacité à travailler dans un cadre interdisciplinaire. Les compétences recherchées sont :

- Modélisation de problèmes découlant des différentes disciplines des science à l'aide d'outils mathématiques et informatiques.
- Compréhension générale des approches modernes des sciences des données et de l'Intelligence Artificielle, y compris pour les modèles moins répandus commes les modèles de Kolmogorov-Arnold.
- Curiosité d'attaquer un problème de recherche interdisciplinaire dans un contexte international et ouverture d'esprit pour la découverte de la langue anglaise utilisée dans un contexte universitaire de recherche.
- Compétences basiques en programmation et en génie logiciel.

La curiosité scientifique, l'autonomie, la rigueur et la capacité à analyser des résultats expérimentaux seront particulièrement appréciées.

Modalités de candidature

Les candidatures doivent être envoyées à Christoph Lauter (cqlauter@utep.edu), Martine Ceberio (mceberio@utep.edu), Olga Lauter (olauter@utep.edu) et Hanane Benmaghnia (hanane.benmaghnia@lip6.fr). Elles doivent inclure :

- un curriculum vitae,
- une lettre de motivation,
- des liens vers des projets personnels, professionnels ou universitaires représentatifs, par exemple des dépôts Git (si disponibles) ou publications antérieures.

References

[1] Abigail Galarza, Valeria Quintero, Christoph Lauter, Olga Lauter, Alex Mayer, and Benjamin Good. Exploring ai-assisted translation and climate data extraction from alaskan russian missionary diaries: Arctic climate research. In *2026 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, pages 65–68, 2026.