



Offre n°2024-08393

## Stagiaire de la recherche : Verbalisation des Comportements Sociaux Audiovisuels pour le Dialogue Multimodal

Type de contrat : Convention de stage

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Stagiaire de la recherche

### Contexte et atouts du poste

Ce stage s'inscrit dans le cadre du projet ECOBOT associant la Faculté de Médecine, l'association étudiante ECN, le Laboratoire lorrain de recherche en informatique et ses applications (LORIA) et le Laboratoire interuniversitaire des sciences de l'éducation et de la communication (LISEC).

L'objectif est de produire un système de verbalisation textuelle des comportements sociaux audiovisuels d'un étudiant en médecine dans le contexte des examens pratiques ECOS.

Le/la stagiaire sera co-encadré(e) par Tom Bourgeade (Post-doctorant, bourse Lorraine Université d'Excellence) et Emmanuel Vincent (Directeur de Recherche, Inria). Il/elle aura accès aux ressources de calcul du LORIA et aura l'opportunité d'interagir avec les autres membres du projet au LORIA et en dehors.

### Mission confiée

Dans le cadre des Examens Cliniques Objectifs Structurés (ECOS) pour les étudiants en sixième année de médecine, le projet ECOBOT vise à simuler des patients et des examinateurs virtuels afin de permettre un accès plus large à cet exercice pratique. Lors des ECOS, les étudiants participent à des scénarios de sept minutes, représentant diverses interactions qu'un médecin peut avoir avec un patient, telles que des entretiens médicaux, des examens cliniques, l'annonce d'un résultat, etc. Le développement de patients et d'examineurs virtuels repose sur des grands modèles de langue (Large Language Models, LLMs) locaux pour des raisons de latence et de protection des données. La formation aux compétences de communication associées, telles que l'empathie, nécessite que ces LLMs prennent en compte non seulement la modalité textuelle du dialogue mais aussi les comportements sociaux (hausser le ton, sourire, montrer une zone douloureuse, etc.) véhiculés par les modalités vocale, faciale et gestuelle. Cependant, [les LLM multimodaux locaux existants](#) sont souvent limités aux données textuelles et aux images statiques et incapables d'analyser la voix ou les mouvements visuels.

En s'inspirant de travaux tels que [MONAH](#), nous souhaitons explorer une approche par verbalisation, c'est-à-dire par encodage et décodage de la prosodie, des expressions faciales et de la gestuelle de l'étudiant-docteur et du patient virtuel sous forme de didascalies textuelles. Dans ce stage, nous nous intéresserons seulement à la partie encodage. En expérimentant différentes approches et modèles, tels que [OpenFace](#) pour la reconnaissance d'unités d'action et d'expressions du visage, ou bien en utilisant un LLM visuel [1; 2] avec le prompting adéquat, l'objectif sera d'insérer dans le flux narratif de la conversation dont le contenu verbal sera capturé par un module de reconnaissance de la parole tel que [Whisper](#) des didascalies correspondant à une verbalisation des comportements sociaux audiovisuels détectés.

### Principales activités

L'objectif de ce stage est de produire un système de verbalisation des comportements sociaux audiovisuels d'un étudiant en médecine dans le contexte des ECOS, en s'inspirant et en étendant l'approche MONAH à un plus large éventail de comportements. Ceux-ci pourront ensuite être utilisés pour enrichir les transcriptions en entrée des LLMs simulant le patient et l'examineur virtuels du projet ECOBOT.

Les objectifs détaillés sont les suivants :

- Catalogage et détection des comportements sociaux** : Dans un premier temps, un catalogue des comportements sociaux utiles au projet et des méthodes pouvant être utilisées pour les détecter sera établi. Ce travail se basera sur les centaines d'enregistrements vidéo d'ECOS déjà collectés dans le cadre du projet, sur la présélection semi-automatique des enregistrements d'intérêt (à l'aide des retours effectués par les évaluateurs humains par exemple), et sur la discussion avec les médecins membres du projet.
- Verbalisation des expressions** : Afin d'étendre les fonctionnalités des approches comme MONAH

basées sur des motifs et règles fixes, différentes méthodes plus dynamiques seront explorées. Par exemple, en employant un LLM visuel tel que Llava [1; 2] avec une stratégie de prompting adéquate, des instants-clés dans le flux vidéo pourraient être ainsi verbalisés. Différentes solutions pour l'identification de ces instants-clés seront envisagées et évaluées.

3. **Évaluation** : Les systèmes produits seront enfin évalués, à la fois qualitativement et quantitativement, au travers de différentes métriques. Pour cela, une petite partie des données pourra être annotée par les membres du projet ECOBOT. Une démarche d'apprentissage actif sera également considérée.

## Compétences

**Compétences techniques et niveau requis :**

- Très bon niveau en informatique attendu (programmation Python) ;
- Connaissances et compétences requises en Apprentissage Automatique, de préférence (dans cet ordre) en Traitement Automatique des Langues (notamment pour le français), en Reconnaissance Automatique de la Parole et Traitement du Signal, en Vision par Ordinateur, et/ou en Reconnaissance des Expressions Faciales et des Gestuelles.

**Langues** : Français et Anglais

## Avantages

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Sécurité sociale

## Rémunération

4.35 €/heure

## Informations générales

- **Thème/Domaine** : Langue, parole et audio
- **Ville** : Villers lès Nancy
- **Centre Inria** : [Centre Inria de l'Université de Lorraine](#)
- **Date de prise de fonction souhaitée** : 2025-02-17
- **Durée de contrat** : 6 mois
- **Date limite pour postuler** : 2024-12-22

## Contacts

- **Équipe Inria** : [MULTISPEECH](#)
- **Recruteur** :  
Bourgeade Tom / [tom.bourgeade@loria.fr](mailto:tom.bourgeade@loria.fr)

## A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

**Attention**: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

## Consignes pour postuler

**Sécurité défense** :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable,

tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

**Politique de recrutement :**

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.