Offre n°2024-07503

PhD Position F/M Explainability of multimodal neural models for the analysis of socio-emotional skills: application to the training of medical students

Le descriptif de l'offre ci-dessous est en Anglais

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Doctorant

Contexte et atouts du poste

Oral communication skills are essential in many situations and have been identified as the core skills of the 21st century. They are now assessed within the curricula themselves by examinations such as the *grand oral du baccalauréat* in high school or objective structured clinical examinations (OSCEs) in medical studies.

This PhD is part of a partnership between Inria and the Medical School of the Université de Lorraine aiming to develop a conversational agent helping medical students to train for OSCEs. During an OSCE, a student faces a "standardized" patient and must ask a series of questions to reach the correct diagnosis. The relevance of the questions asked and the non-verbal communication are graded by an evaluator. Our long-term goal is for the conversational agent to replace both the standardized patient and the evaluator during training sessions, in order to maximize the number of situations seen by students before the actual exam.

The PhD will be co-supervised byEmmanuel Vincent and Chloé Clavel. The PhD student will be based at Inria Nancy to interact with medical students and other staff involved in the project and to benefit from the expertise of the Multispeech team on non-verbal speech attributes and gestures, and spend 12-18 months at Inria Paris to leverage the complementary expertise of the Almanach team on verbal attributes, socio-emotional behaviors and explainability.

Mission confiée

The goals of this PhD are to characterize the communication skills of medical students in the context of OSCE training, to design new explainability approaches providing them with useful feedback, to evaluate the relevance of this feedback, and to contribute to integrating the results into an OSCE training app currently being developed.

This work will be based on a corpus of multimodal data (voice and video of the student and the standardized patient, and questionnaires completed by the student, the standardized patient and the human evaluator after each OSCE) whose collection has begun and will continue throughout the thesis. Each week, 100 to 150 students each take 2 OSCEs lasting 7 minutes, representing a total volume of around 1,000 hours per year that is significantly larger than existing corpora for the study of socio-emotional behaviors.

Principales activités

The expression of socio-emotional skills will be analyzed by extracting the textual transcription and a set of low-level audio and visual features using standard libraries and by deriving high-level features using a neural network trained in a self-supervised manner by predicting the features of the next speech turn [1,2] and/or in a supervised manner by predicting the score and comments assigned by the evaluator. This approach could be supplemented by supervised learning on other corpora annotated in socio-emotional skills such as POM or the MT180 corpus annotated in persuasion capacity [3]. The questionnaires may be gradually modified in order to collect precise comments on the students' performance.

Explainability methods [4,5] will then be devised to identify the high-level multimodal features linked to each socio-emotional skill. We will evaluate the post-hoc explainability method explored in our previous work [6] to identify important time instants and social signals in job interviews. We will also consider counterfactual reasoning [7], a state-of-the-art explainability method which has not yet been explored in this context. The explanations thus found will be verbalized in textual form using a large language model (LLM). The use of Chain-of-Thought LLM architectures capable of generating textual explanations will also be considered [8]. Experimental protocols will be developed to evaluate the explanations provided and in particular assess their ability to provide relevant feedback to the students.

[1] M. McNeill & R. Levitan, “An autoregressive conversational dynamics model for dialogue systems”, in Interspeech, 2023.
[2] E. Chapuis, P. Colombo, M. Manica, M. Labeau & C. Clavel, “Hierarchical pre-training for sequence labelling in spoken dialog”, in Findings of EMNLP, 2020.
[3] A. Barkar, M. Chollet, B. Biancardi & C. Clavel, “Insights into the importance of linguistic textual features on the persuasiveness of public speaking”, in ICMI, 2023.
[4] L.H. Gilpin, D. Bau, B.Z. Yuan, A. Bajwa, M. Specter & L. Kagal, “Explaining explanations: An overview of interpretability of machine learning”, in DSAA, 2018.
[5] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova & C. Zhong, “Interpretable machine learning: Fundamental principles and 10 grand challenges”, Statistic Surveys, 2022.
[6] L. Hemamou, A. Guillon, J.C. Martin & C. Clavel, “Multimodal hierarchical attention neural network: Looking for candidates behaviour which impact recruiter's decision”, IEEE Transactions on Affective Computing, 2021.
[7] S. Verma, J. Dickerson & K. Hines, “Counterfactual explanations for machine learning: Challenges revisited”, in CHI HCXAI workshop, 2021.
[8] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le & D. Zhou, “Chain-of-Thought prompting elicits reasoning in large language models”, arXiv:2201.11903, 2022.

Compétences

MSc degree in speech processing, NLP, computer vision, machine learning, or in a related field.
Strong programming skills in Python/Pytorch.
Prior experience with speech, text and/or video processing is an asset.
French speaking skills are a plus.

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération

2100€ gross/month the 1st year

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 :** 2024-10-01
- **Durée de contrat :** 3 ans
- **Date limite pour postuler :** 2024-05-12

Contacts

- **Équipe Inria :** MULTISPEECH
- **Directeur de thèse :**
  Vincent Emmanuel / emmanuel.vincent@inria.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.