# Offer #2025-08731

# PhD Position F/M pre-doc position / Deep Neural Networks for Analyzing Non-Verbal Behavior during Clinical Interactions

**Contract type :** Fixed-term contract

**Level of qualifications required :** Graduate degree or equivalent

**Fonction :** Tempary Research Position

## About the research centre or Inria department

The Inria centre at Université Côte d'Azur includes 42 research teams and 9 support services. The center's staff (about 500 people) is made up of scientists of di?erent nationalities, engineers, technicians and administrative staff. The teams are mainly located on the university campuses of Sophia Antipolis and Nice as well as Montpellier, in close collaboration with research and higher education laboratories and establishments (Université Côte d'Azur, CNRS, INRAE, INSERM ...), but also with the regional economic players.

With a presence in the fields of computational neuroscience and biology, data science and modeling, software engineering and certification, as well as collaborative robotics, the Inria Centre at Université Côte d'Azur  is a major player in terms of scientific excellence through its results and collaborations at both European and international levels.

## Context

Inria, the French National Institute for computer science and applied mathematics, promotes "scientific excellence for technology transfer and society". Graduates from

the world's top universities, Inria's 2,700 employees rise to the challenges of digital sciences. With its open, agile model, Inria is able to explore original approaches with its partners in industry and academia and provide an efficient response to the multidisciplinary and application challenges of the digital transformation. Inria is the source of many innovations that add value and create jobs.

**Team**

The STARS research team combines advanced theory with cutting edge practice focusing on cognitive vision systems.

*Team web site : [https://team.inria.fr/stars/](https://team.inria.fr/stars/)*

# Assignment

The Inria STARS team is seeking for a pre-doc researcher with strong background in computer vision, deep learning and machine learning.

"Actions speak louder than words". Humans are complex beings, and they often convey a wealth of information not through their words but through their actions and demeanor. Non-verbal behaviors can offer crucial insights into their emotional state, pain level, or anxiety, often more eloquently than words alone. The analysis of non-verbal communication is of critical importance in the diagnostic landscape. Decoding non-verbal cues in a clinical setting requires healthcare professionals to be astute observers, picking up on nuances that may be subtle yet critical. The challenge lies in accurately interpreting these cues, as they can vary greatly from one individual to another.

To address this challenge, automated systems capable to detect non-verbal behaviors and their corresponding meanings can assist healthcare providers. Such technology is not to replace medical experts but rather to act as their supportive tool.

The primary objective of this technical internship is to lead the development of an advanced AI model for Human Behavior Understanding to identify non-verbal cues expressed by patients, and then interpreting the cues to derive critical insights about their health. Traditionally, computer vision methodologies encompassing skin color analysis, shape analysis, pixel intensity examination, and anisotropic diffusion were used to identify body parts and trace their activities. However, these algorithms provided limited flexibility because of their domain-specific nature. Deep learning methods can be used to deal with this issue as they offer more training flexibility, and better performance results. The overarching goal is to provide a real-time, data-driven analysis of non-verbal cues exhibited by patients during clinical interactions, thereby delivering invaluable insights to healthcare practitioners.

# Main activities

With our vision of evidence-based diagnosis, we will develop explainable methods for biomarker detection from audiovisual and physiological data. Generally, AI models are based on machine learning concepts that find intrinsic correlations between multiple input channels and the true labels. To be able to model the complex action patterns, we need to go beyond deep learning by incorporating some semantic modeling within the deep learning pipeline, which today consists of a combination of CNN and transformers. These complex action patterns include composite actions and concurrent actions occurring in long untrimmed videos. Existing methods have mostly focused on modeling the variation of visual cues across time locally or globally within a video. However, these methods consider the temporal information without any further semantics. Videos may contain rich semantic information such as objects, actions, and scenes. Real-world videos contain also many complex actions with inherent relationships between action classes at the same time steps or across distant time steps. Modeling such class-temporal relationships can be extremely useful for locating actions in those videos. Therefore, semantic relational reasoning can help determine the action instance occurrences and locate the actions in the video, especially for complex actions in the video.

Going beyond classical deep CNNs, our first attempts will be to extract the relevant semantics using large language-vision models (LVMs). However, large foundation models work really well and have almost pixel-level attention, although they are not scalable. Their monstrous size makes it hard to fine-tune. What we will do is instead of learning temporal relations from scratch we will exploit the optical flow of attention maps and its information of motion on a feature level, which does not require much processing to classify actions. This optical flow is obtained using the attention maps from processed frames of videos using image foundation models. Adapters have shown to work well and provide a downsampled embedding of the hidden layers of the base model which is easy to work with. We intend to move towards the direction of designing plugin architectures that makes large transformer models more efficient by omitting fine-tuning of the whole models and other additions.

# Skills

Candidates must hold a Master degree or equivalent in Computer Science or a closely related discipline by the start date.

The candidate must be grounded in the basics of computer vision, have solid mathematical and programming skills.

Preferably in Python, OpenCV, deep learning framework Pytorch or Tensorflow.

The candidate must be committed to scientific research and strong publications.

# Benefits package

- 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 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
- Contribution to mutual insurance (subject to conditions)

# Remuneration

Gross Salary per month: 2200€ brut per month

# General Information

- **Theme/Domain :** Vision, perception and multimedia interpretation
  Biologie et santé, Sciences de la vie et de la terre (BAP A)
- **Town/city :** Sophia Antipolis
- **Inria Center :** Centre Inria d'Université Côte d'Azur
- **Starting date :** 2025-07-01
- **Duration of contract :** 12 months
- **Deadline to apply :** 2025-04-19

# Contacts

- **Inria Team :** STARS
- **Recruiter :**
  Balazia Michal / michal.balazia@inria.fr

# About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

# Instruction to apply

**Defence Security :**

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

**Recruitment Policy :**

As part of its diversity policy, all Inria positions are accessible to people with disabilities.