

Engineer in data and machine learning for film analysis with data-centric AI

General information

Key-words: Statistical data analysis, Deep learning, Film datasets, Feature extraction, Multimodal data, data-centric AI

Function: Engineer

Level of qualifications required: Master/engineering degree

Level of experience: Recently graduated

Location:

Laboratoire I3S, Université Côte d'Azur, CNRS
06900 Sophia Antipolis, France

Starting date: November 2023-January 2024

Duration of contract: 18 months

Instruction to apply: Send to below email addresses:

Required:

CV detailing prior education, project experience and results

Motivation letter (*it is strongly recommended to address the context and assignment of the post, and not just the competences*)

Contacts for 3 references

Master's transcripts

Appreciated: personal website, project code repositories

Any incomplete application will be automatically discarded.

Supervisor:

Prof. Lucile Sassatelli, Full Professor, Fellow of IUF

lucile.sassatelli@univ-cotedazur.fr, <https://www.i3s.univ-cotedazur.fr/~sassatelli>

Mission

The engineer will be recruited in the Signal Image and Systems department of the I3S laboratory, and given tasks related to film analysis with AI (machine learning - ML) models. They will be tasked with providing the necessary components to produce and publish new datasets to train ML models. The datasets will be prepared under Open Science guidelines and submitted to most prestigious venues such as the NeurIPS Dataset track. These components will enable the development of data-centric AI approaches.

The main tasks will involve:

- the pre-processing workflow for the different modalities (video, text, audio),
- advanced statistical analyses of the datasets,
- testing and result analysis of several deep learning baselines for the target tasks of the project,

- precise documentation of the data (such as *Datasheets for datasets* or *Dataset nutrition scores*).

Activities

Main activities:

- establish and manage data (image, text, metadata) processing workflow
- testing and integrating existing code repositories or libraries
- seek and propose technical solutions to data processing needs
- formatting data for machine learning algorithms
- Python code development
- writing technical documentation and reports

Additional activities

- participating in project meetings
- assist PhD students and project partners in their regular needs of off-the-shelf AI tool testing

Pre-requisites

Required

- Excellent programming skills in Python,
- Solid experience with machine learning libraries (scikit learn, tensorflow, torch) and algorithms,
- Strong competence and prior experience with computer vision, image processing, text processing tools and libraries (e.g., opencv, ffmpeg, cimg, dlib, nltk, Lemur),
- Good familiarity with data processing workflows for machine learning,
- Expertise in code repository management (git),
- Good engineering skills to debug and compile repositories with source code in C++ and Matlab,
- Seeking, understanding, using, and writing technical documentation,

Appreciated

- Knowledge of virtualization and software packaging (e.g., Docker, Singularity),
- A good level of written and spoken French and English

Work context

The TRACTIVE project is a 4-year research project funded by the French National Research Agency (ANR) and involving 6 academic laboratories. The scientific objective of TRACTIVE is to characterize and quantify gender representation and women objectification in films and visual media by designing an AI-driven multimodal (visual and textual) discourse analysis. To address this objective, TRACTIVE brings together researchers from computer science, media studies, linguistics, and gender studies.

In addition to the permanent researchers, several non-permanent staff members are already working on the project, including PhD students and an engineer.

Remuneration

Gross salary: from 2200€ /month depending on experience

French social security included