Job Opening - Internship In Machine Learning

Teaching machine fairness

As part of the ANR Diké project: Biases of compressed language models, which brings together the company NaverLab and the laboratories ERIC and Hubert Curien, we are recruiting a master degree intern for a duration of 4 to 6 months.

Project Background:

Natural Language Processing (NLP), a subfield of Artificial Intelligence (AI), aims to automate the processing of written text, covering tasks such as text analysis and generation (e.g., automatic translation). Deep learning, particularly the "transformer" architecture found in the latest language models (ChatGPT, LLama, etc.), plays a crucial role in modern NLP. However, these models were shown to be biased by several studies, including deep racial and gender stereotypes. The Diké Project focuses on Large Language Model fairness and ethic, and the impact of compression on these biases.

Expected Work:

The intern will work closely with the Diké project team to conduct advanced research on fairness of Large Language Models. The results of this research will contribute to the creation of more ethical language models, paving the way for broader and fairer applications of AI in natural language processing.

This internship will focus on a novel subject: teaching machine fairness. While most existing works try to make the model forget the sensitive attribute (e.g. gender), the objective here is to find a way to teach the model what is fair and what isn't in order to be able to more accurately control its decision.

The recruited intern will continue the work already started in this direction by the researchers of the project.

Required Skills:

The candidate must possess strong skills in Machine Learning (model design, proficiency in deep learning frameworks such as PyTorch/TensorFlow). Additionally, they should have advanced skills in Python, a strong affinity for textual data and Large Language Models (GPT, LLama, PaLM), and their application (notably via HuggingFace).

Salary:

4.05 euros per hour, in accordance with university salary scales

Additional Information:

This internship is a co-supervision between SAint-Etienne and Lyon, even if the recruited intern will be physically in Saint-Etienne. Some visits to the Lyon lab will be planned through the internship.

The Laboratoire Hubert Curien (https://laboratoirehubertcurien.univ-st-etienne.fr) is a joint research unit (UMR 5516) of the Université Jean Monnet de Saint-Etienne, the Centre National de la Recherche Scientifique (CNRS) and the Institut d'Optique Graduate School. The laboratory's research activities are organized into two scientific departments: Optics, Photonics and Surfaces, and Computing, Security and Imaging. The Data Intelligence team, in which the new recruit will work, specializes in Machine Learning.

The ERIC laboratory (a research unit of the Universities of Lyon 2 and Lyon 1) develops theoretical and applied research in the fields of data science and business intelligence. It aims to make the most of large, complex databases, particularly in the fields of literature, languages, humanities and social sciences (LLSHS), in the areas of Data Science, Business Intelligence and Digital Humanities.

The recruited person will have access to a workstation with a computer enabling the use of the laboratory's computing cluster.

To apply, please send a detailed CV and a letter of motivation to antoine.gourru@univ-st-etienne.fr and julien.velcin@univ-lyon2.