Biases of language models for low resource fine tuning

Post-Doc ou Ingénieur de Recherche
6 mois à 2 ans

----------English Below----------

Dans le cadre du projet ANR **Diké : Biases of compressed language models,** réunissant l'entreprise NaverLab et les laboratoire ERIC et Hubert Curien, nous recrutons un.e post-doctorant.e ou ingénieur.e de recherche pour une durée de 6 mois à 2 ans

Contexte du projet
Le Traitement du Langage Naturel (NLP), un sous-domaine de l'Intelligence Artificielle (IA), vise à automatiser le traitement du texte écrit, couvrant des tâches telles que l'analyse et la génération de texte (par exemple, la traduction automatique). L'apprentissage profond, en particulier l'architecture "transformer" présente dans les modèles de langage les plus récents (ChatGPT, LLama, etc) joue un rôle essentiel dans le NLP moderne. Ces modèles, avec leur nombre croissant de paramètres, rencontrent des défis lors du déploiement en raison de leur taille. Les chercheurs y remédient avec des techniques de compression telles que la quantification des poids, l'élagage du modèle et la distillation du modèle, permettant d'obtenir des modèles plus petits sans perte significative de précision. Le projet Diké se concentre sur l'étude des effets de compression en NLP, notamment les biais et l'éthique, dans le but de créer des techniques de compression plus équitables/éthiques.

Travail attendu
Le.a chercheur.se devra travailler en étroite collaboration avec l'équipe du projet Diké pour mener des recherches avancées sur les techniques de compression en NLP. Les résultats de ces recherches aideront à créer des modèles de langage plus compacts et plus éthiques, ouvrant ainsi la voie à des applications plus larges et plus équitables de l'IA dans le traitement du langage naturel. Plus spécifiquement les verrous scientifiques identifiés sont :

- Implémenter les méthodes récentes de sur apprentissage de modèles de langue large, plus précisément les approches faibles ressources (“low-resources”), telles que Lora, Qlora, le Prompt tuning.
- Évaluer les biais de ces nouvelles approches, par exemple sur les jeux de données classiques du domaine comme BiasInBios.
- Etudier egalement la robustesse de ces approches, notamment dans le cadre de données déséquilibrées
- Proposer des approches de fine-tuning à faibles ressources plus équitable et éthiques

Compétences requises
Le/la candidat.e doit posséder des compétences solides en Apprentissage Automatique (conception de modèles, maîtrise des framework d'apprentissage deep tels que PyTorch/TensorFlow), mais aussi des compétences avancées en Python, une forte appétence pour les données textuelles, le question answering et les Modèles de Langues dits Larges (GPT, LLama, PaLM), ainsi que le surapprentissage et l'application de ces derniers (Notamment via HuggingFace).

Salaire

Environ 1 745 euros net pour un.e titulaire de master, 2303 euros net pour un.e titulaire de doctorat.

Informations supplémentaires
Le lieu d'accueil est le Laboratoire Hubert Curien (https://laboratoirehubertcurien.univ-st-etienne.fr), unité mixte de recherche (UMR 5516) de l'Université Jean Monnet de Saint-Etienne, du Centre National de la Recherche Scientifique (CNRS) et de l'Institut d'Optique Graduate School. Il est composé d'environ 90 chercheurs, professeurs et maîtres de conférences, 20 ingénieurs et personnels administratifs et 130 doctorants et post-doctorants. Nos activités de recherche sont organisées selon deux départements scientifiques : Optique, photonique et surfaces et Informatique, sécurité, image. L'équipe Data Intelligence, au sein de laquelle la personne recrutée travaillera, est spécialisée dans le domaine du Machine Learning

Le salaire est modulable en fonction de l'expérience du/de la candidat.e. La personne recrutée aura accès à un poste de travail avec un ordinateur permettant l'utilisation du cluster de calcul du laboratoire. Le début du contrat est prévu pour début Janvier 2023.

Pour candidater, merci d'envoyer à antoine.gourru@univ-st-etienne.fr et julien.velcin@univ-lyon2. : un CV détaillé et une lettre de motivation, tout cela **avant le**

----------English Version ----------

Job Opening - Postdoctoral Researcher or Research Engineer

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 postdoctoral researcher or research engineer for a duration of 6 months to 2 years.

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, with their increasing number of parameters, face challenges during deployment due to their size. Researchers address this by using compression techniques such as weight quantization, model pruning, and model distillation, which allow obtaining smaller models without significant loss of accuracy. The Diké project focuses on studying the effects of compression in NLP, particularly biases and ethics, with the aim of creating more equitable/ethical compression techniques.

Expected Work:

The researcher will work closely with the Diké project team to conduct advanced research on compression techniques in NLP. The results of this research will contribute to the creation of more compact and ethical language models, paving the way for broader and fairer applications of AI in natural language processing. Specifically, the identified scientific challenges are as follows:

- Implement recent methods of fine-tuning large language models, particularly low-resource approaches such as Lora, Qlora, and Prompt tuning.
- Evaluate the biases of these new approaches, for example, on classic domain datasets like BiasInBios.
- Study the robustness of these approaches, particularly in the context of imbalanced data.
- Propose fairer and ethical low-resource fine-tuning approaches.

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, question answering, and Large Language Models (GPT, LLama, PaLM), as well as fine-tuning and their application (notably via HuggingFace).

Salary:

Approximately 1,745 euros net for a Master's degree holder, 2,303 euros net for a PhD holder.

Additional Information:

The position will be based at the Hubert Curien Laboratory (https://laboratoirehubertcurien.univ-st-etienne.fr), a joint research unit (UMR 5516) of the Jean Monnet University of Saint-Etienne, the National Center for Scientific Research (CNRS), and the Graduate School of Optics Institute. The laboratory consists of approximately 90 researchers, professors, and associate professors, 20 engineers and

administrative staff, and 130 doctoral and postdoctoral researchers. Research activities are organized into two scientific departments: Optics, Photonics, and Surfaces, and Computer Science, Security, and Imaging. The Data Intelligence team, where the recruited person will work, specializes in the field of Machine Learning.

The salary is negotiable based on the candidate's experience. The recruited person will have access to a workstation with a computer enabling the use of the laboratory's computing cluster. The contract is expected to start in early January 2023.To apply, please send a detailed CV and a letter of motivation to antoine.gourru@univ-st-etienne.fr and julien.velcin@univ-lyon2