Offre n°2023-06862

PhD Position F/M Guaranteed Fairness in Machine Learning

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

Type de contrat : CDD

Niveau de diplôme exigé : Bac + 5 ou équivalent

Fonction : Doctorant

A propos du centre ou de la direction fonctionnelle

The Inria University of Lille centre, created in 2008, employs 360 people including 305 scientists in 15 research teams. Recognised for its strong involvement in the socio-economic development of the Hauts-de-France region, the Inria University of Lille centre pursues a close relationship with large companies and SMEs. By promoting synergies between researchers and industrialists, Inria participates in the transfer of skills and expertise in digital technologies and provides access to the best European and international research for the benefit of innovation and companies, particularly in the region.

For more than 10 years, the Inria University of Lille centre has been located at the heart of Lille's university and scientific ecosystem, as well as at the heart of Frenchtech, with a technology showroom based on Avenue de Bretagne in Lille, on the EuraTechnologies site of economic excellence dedicated to information and communication technologies (ICT).

Contexte et atouts du poste

The selected PhD candidate will be based in Lille in the MAGNET team. The main objective of the team is to develop ethically acceptable machine learning algorithms focusing on fairness, privacy, and decentralized learning and to empower end users of artificial intelligence. The PhD candidate will be under the supervision of Michaël Perrot and Marc Tommasi.

Mission confiée

Machine learning is nowadays used in various applications, such as medical diagnosis and speech recognition. Its success stems from the performance of learned models, sometimes reaching human-level capabilities. However, deploying these models on a larger scale requires more than just accuracy, and it is imperative to consider fairness and trustworthiness when human lives are affected. For instance, a model used for medical diagnosis or an automated hiring process should not be biased against subgroups of the population. A plethora of approaches have been proposed in the scientific literature to overcome such issues by training models to maintain reasonable levels of accuracy while limiting discrimination of individuals or subgroups.

The effectiveness of these approaches is often assessed through empirical investigation on a limited range of datasets. Only a small number of methods offer theoretical guarantees of the fairness of the learned models, either in training or in generalization. In addition, these guarantees are not always in line with the settings where the models are applied. For example, they may solely consider stochastic decisions and disregard the need of deterministic predictions in practical applications. The aim of this PhD is to bridge this gap by proposing novel frameworks to learn fair models. For instance, potential avenues could include learning deterministic predictors from stochastic outcomes or introducing new algorithms with inherent fairness guarantees.

Fairness is not the sole constraint that should be enforced to obtain a trustworthy model. Other essential properties appear in the literature, among which privacy stands out as a very important requirement. The underlying idea is that the learned models should not leak information about the data that was used to train them. Although fairness and privacy have each been extensively examined separately, only recently have their interactions garnered more attention. A secondary goal of this thesis is to consider the interplay between these two notions, in particular in the context of the frameworks proposed to learn fair models.

Principales activités

1. Review and follow the existing literature on Fair Machine Learning with a particular focus on well founded methods.
2. Propose new fairness frameworks to bridge the gap between stochastic models and the derived deterministic decisions that are used in practice.
3. Study the problem of generalization in fair machine learning.
4. Investigate the interplay between fairness, privacy, and utility from a theoretical, empirical, and algorithmic point of view.
5. Publish and present results in top machine learning conferences and journals.

**Compétences**

A good candidate will have the following skills:

- A good command of English
- A strong background in mathematics
- A good knowledge of machine learning, statistics and algorithms
- Some experience with implementation and experimentation
- Some knowledge on fairness would be a plus

Please follow the instructions given in [https://team.inria.fr/magnet/how-to-apply/](https://team.inria.fr/magnet/how-to-apply/) to set up your application file.

**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 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**

2082€ gross/month and 2190€ gross/month for the final year

**Informations générales**

- **Thème/Domaine**: Optimisation, apprentissage et méthodes statistiques
  Statistiques (Big data) (BAP E)
- **Ville**: Villeneuve d'Ascq
- **Centre Inria**: Centre Inria de l'Université de Lille
- **Date de prise de fonction souhaitée**: 2024-10-01
- **Durée de contrat**: 3 ans
- **Date limite pour postuler**: 2023-11-30

**Contacts**

- **Équipe Inria**: MAGNET
- **Directeur de thèse**: michael.perrot@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.

**L'essentiel pour réussir**

A successful candidate will

- Collaborate in the team and where applicable with external researchers and engineers
- Organize work efficiently and make a good balance between the several priorities
- Report regularly

**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**
Resume and cover letter

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.