Offer #2024-07711

PhD Position F/M Fairness and Privacy in Machine Learning

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

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).

Context

The selected PhD candidate will be based in Lille in theMAGNET 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.

Assignment

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 privacy when human lives are affected. For instance, a model used for medical diagnosis should not be biased against subgroups of the population. Similarly, a model learned on datasets containing personal informations should not leak it to the public. 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 or preserving privacy. While the performance of these fairness and privacy preserving approaches has been extensively studied in isolation, only a small number of works addressed the problem of their respective impact on one another.

Fairness and privacy may negatively impact one another. Hence, fairness enforcing algorithms often require access to sensitive attributes about specific individuals, such as gender or race, and these sensitive attributes may be leaked by the learned models. Similarly, privacy constraints, that often try to mask specific characteristics of the individuals to prevent leakage, may have a larger impact on smaller subgroups of the population on which the learning algorithms already have difficulties capturing a precise signal. On the other hand, fairness and privacy may complement one another. Hence, the ideal fair models should make predictions that are independent of sensitive attributes, one way to achieve this being to completely hide this information in the data. Similarly, privacy preserving approaches tend to blur the separation between specific examples, making hem indistinguishable from one another and thus promoting fairness by preventing personalized predictions.

The aim of this PhD is to study the interplay between fairness and privacy. Potential research avenues include theoretical studies of the impact that different fairness enforcing and privacy preserving mechanisms may have on one another. Such results could take the form of upper and lower bounds on the trustworthiness levels achieved by specific algorithms or the definition of sufficient conditions under which the impact that these two notions have on one another is limited or controlled.

Main activities

1. Review and follow the existing literature on the interplay between Fairness and Privacy in Machine Learning with a particular focus on theoretical results.
2. Propose theoretical frameworks to quantify the impact that fairness and privacy may have on one

another.
3. Propose new algorithmic solutions to mitigate this impact.
4. Publish and present results in top machine learning conferences and journals.

Skills

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 or privacy would be a plus

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

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
- Social security coverage

General Information

- **Theme/Domain :** Optimization, machine learning and statistical methods
  Statistics (Big data) (BAP E)
- **Town/city :** Villeneuve d'Ascq
- **Inria Center :** Centre Inria de l'Université de Lille
- **Starting date :** 2024-10-01
- **Duration of contract :** 3 years
- **Deadline to apply :** 2024-06-21

Contacts

- **Inria Team :** MAGNET
- **PhD Supervisor :**
  Perrot Michael / michael.perrot@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.

The keys to success

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

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

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.