

## **Chair Data Science and Artificial Intelligence for Digitalized Industry and Services**

### **Internship project**

#### **Subject**

Kernel-based multi-scale learning with the alpha-procedure

#### **Possibility to continue as a PhD candidate**

YES (Funding to be confirmed)

#### **About the chair**

The Chair Data Science and Artificial Intelligence for Digitalized Industry and Services (DSADIS), lead by Florence d'Alché-Buc, a Professor in the department Image, Data, Signal of Telecom Paris, unites five industrial partners: Airbus Defence & Space, Engie, Idemia, Safran et Valeo. It's general objective is to develop, in collaboration with the partners, teaching and research of the international level.

Its four principal research directions are:

1. Building predictive analytics on time series and data streams.
2. Exploiting large scale, heterogeneous, partially labeled data.
3. Machine Learning for trusted and robust decision.
4. Learning through interactions with environment.

#### **Description of the internship**

##### **Supervision**

Stephan Clemencon (<https://perso.telecom-paristech.fr/clemenco/>)

##### **Location and dates of the internship**

Address : Télécom Paris, 19 Place Marguerite Perey, 91120 Palaiseau

Date of the beginning of the internship : beginning April 2024

##### **Team where the thesis will be written**

Department IDS, Team Signal, Statistique et Apprentissage (S2A)

##### **Keywords**

Fairness and bias in AI.

##### **Detailed subject**

AI systems that make significant decisions for humans, regarding for instance credit lending in the banking sector, diagnosis in medicine, biometric identification for security purposes or recidivism prediction in criminal justice should guarantee that they do not penalize certain segments of the populations (defined through e.g. gender, ethnicity, age). Stimulated by the forthcoming regulation (AI Act), appropriate notions of fairness in AI must be designed, depending on the application considered (e.g. facial recognition, recommendation engine, credit scoring), and the type of fairness sought (individual or group fairness, equal opportunity, parity in mistreatment). Current research

essentially focuses on the definition of relevant metrics for fairness evaluation purpose in various situations and the elaboration of ML methods producing predictive rules meeting fairness constraints (e.g. by modifying a posteriori a pre-trained decision function, incorporating the constraints in the loss function during the learning stage).

Truth should be told, in most approaches currently developed, fairness is generally obtained to the detriment of the accuracy of AI systems. This suggests future lines of research, considering the definition of fairness in conjunction with other merits: explainability, robustness, uncertainty and performance compatibility. Not only must the fairness metrics reflect the societal demand for fairness in an explainable/interpretable manner, which is a challenge in itself in many situations, but they must also lend themselves to reliable/trustworthy statistical estimation and numerical optimization, so as to force the AI rules to meet fairness constraints without damaging seriously predictive performance whenever this is possible. Hence, increasing the specifications of AI systems in this direction, essential to their acceptability, is making demands on the research community and it is the main purpose of this internship to bring answers, relying on methodological expertise and a deep understanding of what is at stake in the applications. Such answers could be the key of a future efficient regulation that would not restrain innovation.

As the subject is vast and ambitious, the intern may possibly start with prolongating the preliminary approaches considered in the references listed below

### **Candidate profile**

Student having master 2 research

- Statistical learning, bases of probability
- Good level of programming (R, C/C++, Python)
- Good command of English

### **Application**

To send on [pavlo.mozharovskyi@telecom-paris.fr](mailto:pavlo.mozharovskyi@telecom-paris.fr):

- Curriculum Vitae
- Personalized motivation letter that explains interest of the candidate in the subject (can be directly in the body of the email)
- Grade reports for recent years
- Contact of a person willing to give recommendation

Incomplete applications will not be considered.

### **References**

[1] Mitigating Gender Bias in Face Recognition Using the von Mises-Fisher Mixture Model. J.R. Conti, S. Cléménçon, N. Noiry, S. Gentric & V. Despiegel. In the **Proceedings of ICML**, 2022.

[2] Learning Fair Scoring Functions: Bipartite Ranking under ROC-based Fairness Constraints. With A. Bellet (INRIA) and R. Vogel (Telecom Paris, IDEMIA). In the **Proceedings of AISTATS**, 2021.