

## Intervenants séance 4

Cette séance, intitulée *L'IA est-elle discriminante : sexiste, raciste, inégalitaire ?* sera organisée en trois parties :

**1<sup>ère</sup> partie : Patrice Bertail** est professeur de Mathématiques (spécialité statistique et probabilité) à l'université Paris Nanterre après avoir été chercheur à l'INRA(E). Il est actuellement directeur du laboratoire MODAL'X UMR CNRS 9023.

### **Biais en apprentissage statistique : une revue et une approche semi-paramétrique par calibration**

Résumé:

Dans cette présentation, nous passons d'abord en revue quelques types de biais pouvant apparaître fréquemment dans des problèmes d'apprentissage statistique : biais de données, biais de sélection, biais d'endogénéité, biais de modèle dans les cas non-stationnaires, biais cognitifs etc... Ces types de biais sont bien connus dans la littérature statistique et plus encore en économétrie mais pas toujours pris en compte en apprentissage. Ces problèmes de biais posent non seulement des problèmes techniques mais aussi éthiques comme nous le verrons sur quelques exemples. Les techniques de corrections de biais sont souvent inspirées de technique de sondage et reposent sur des pondérations adéquates des individus : Vardi dans les années 80's a même donné des conditions nécessaires et suffisantes pour pouvoir corriger de biais (essentiellement de sélection ou d'endogénéité) en présence d'information transversales et/ou marginales. Nous nous intéressons dans une seconde partie au problème de l'apprentissage par transfert ("Transfert Learning") et montrons comment l'incorporation d'un modèle semi-paramétrique permet de corriger des biais à partir de l'observation de quelques moments : l'idée est très proche de la calibration ou du calage sur marge en sondage, les poids permettant de correctement repondérer les risques.

Travail joint avec S. Cléménçon (Télécom, Institut Polytechnique), D. Bounie ((Télécom, Institut Polytechnique), Y. Guyonvarch (INRAE), N. Noiry (Owkin), P. Waelbroeck Télécom, Institut Polytechnique).

**2<sup>ème</sup> partie : Pauline Gourlet**, Designer et chercheuse en sciences sociales, Pauline Gourlet travaille sur les technologies numériques : à travers des enquêtes participatives, elle tente de pluraliser et de politiser les opérations de documentation, de classification, de calcul.

### **Racisme, sexisme et inégalités : IA et dépossession**

Résumé :

Poser la question des "bonnes manières" de calculer pour prévenir toute forme de discrimination est évidemment nécessaire et ce problème a toute sa place dans les conversations sur les développements de l'IA. Mais cette critique interne du système technique se fait souvent au détriment de — même, tend à rendre inaudible — une autre conversation : celle des conditions d'existence des techniques computationnelles. L'enjeu de cette présentation sera de montrer en quoi les développements de l'IA reposent sur des mécanismes de dépossession et perpétuent une organisation sociale raciste, sexiste et largement inégalitaire. Inspirée notamment par les théories féministes de la reproduction sociale, je décrirai trois opérations pour décaler les problèmes de "l'IA", avec un objectif proprement *génératif* : celui de réouvrir des options de développement technique qui se soucient du renouvellement du monde et de la justice.

**3<sup>ème</sup> partie : Débat.**

Les présentations de **Patrice Bertail** et **Pauline Gourlet** seront suivies d'un débat animé par **Hélène Gispert**, Université Paris Saclay, **Samuel Roturier**, UVSQ et **Julien Gargani**, Directeur du Centre d'Alembert.