AI bias and Data Privacy

Counteracting prejudices of machine learning algorithms and managing data privacy
Friday 15 January 2021, 8.45 to 11:15 a.m.

For the past 10 years, machine learning and AI have become increasingly common in high-stakes applications such as healthcare, robotic surgery, autonomous vehicles etc. The reliability of modern algorithms and architectures makes them appealing in a growing number of applications. The ethical implications of pervasive AI-based solutions have already begun to affect EU countries so that the development of ethics, standards, and regulation is emerging as a crucial area of research investigation. In this first webinar bringing together Hi! Paris researchers, corporate donors and public policy makers, we focus on two key ethical issues: the management of AI bias and data privacy.

The explanation of AI bias is often reduced to discussions about bias in training data. The reality is much more complex as bias can be caused by many other factors in all learning processes for real-life applications. First, the algorithm can “hack” the problem by fostering prejudicious behaviors to optimize a cost function which does not assume enough constraints on the input data. Second, the collected data may also be unrepresentative of reality, in which case an algorithm will reflect existing prejudices instead of exposing and solving them. Finally, bias may be introduced during data featuring and processing (e.g., using selection procedures of the input variables).

In the first two research presentations, Christophe Pérignon (HEC Paris) and Stéphan Clémençon (Télécom Paris), will share insights from their work on AI bias. Algorithms have also been under increasing criticism for breaching data privacy and Ruslan Momot (HEC Paris) and Catuscia Palamidessi (Ecole Polytechnique) will present works pinpointing concrete best practices allowing companies to more efficiently and effectively manage data privacy and individual fairness.

In addition to research presentations, this webinar will allow experts in AI and machine learning to join the discussion, in order to detail challenges related to AI bias and data privacy and highlight solutions to promote best practices with respect to fairness and transparency of AI systems.

8.45-9.55– FOUR ILLUSTRATIONS OF RESEARCH COOKING AT HI! PARIS

8.45-8.55 a.m. — Fresh Hi! Paris news.
Gaēl Richard - Télécom Paris
Executive Director of Hi! Paris

8.55-9.10 a.m. — Illustration 1:
Christophe Pérignon — HEC Paris
Associate Dean for Research and Professor of Finance

Title: Algorithmic Fairness in Credit Scoring Models

Teaser:
AI can systematically treat unfavorably a group of individuals sharing a protected attribute (e.g. gender, race, religion). In credit scoring applications, this lack of fairness can severely distort access to credit and expose AI-enabled financial institutions to legal and reputational risks. We develop a unified framework assessing the fairness of AI algorithms and illustrate its efficiency using a dataset of consumer loans.

9.10-9.25 a.m. — Illustration 2:
Stéphan Clémençon — Télécom Paris
Professor of machine learning and head of the research team S2A

Title: On Selection Bias and Fairness Issues in Machine-Learning

Teaser:
In the Big Data era, the exploding quantity of observations available for training predictive rules by means of machine learning techniques is a sort of frequentist nirvana, as promised by the law of large numbers. However, the success of predictive algorithms cannot be not guaranteed by the massiveness of the information at disposal solely, controlling rigorously the conditions for acquiring data and the fairness constraints making the automated rules, once deployed, acceptable by the general public turns out to be essential.

9.25-9.40 a.m. — Illustration 3:
Ruslan Momot — HEC Paris
Assistant Professor of Operations Management

Title: Consumer Privacy Preservation – Firm’s and Regulator’s Perspectives.

Teaser:
With online shopping, loyalty programs, smart devices and many other aspects of business and daily lives, companies collect vast amounts of consumer data. The risk is that these data may be leaked or misused. What are the measures that both companies and regulators can undertake to preserve consumer privacy?

9.40-9.55 a.m. — Illustration 4:
Catuscia Palamidessi — Ecole Polytechnique / INRIA
Director of Research at INRIA

Title: Privacy concerns in the era of machine learning

Teaser:
Machine learning technology is an invaluable asset to social and economical progress, but it has exacerbated the risks of privacy violations and unfair decisions. In this talk, we showcase the issues at stake and the state-of-the-art approaches proposed by the scientific community to mitigate these threats.

9.55 – 10.10 – AN OVERVIEW OF CNIL PRIORITIES

Bertrand Pailhès,
Head of innovation, CNIL.
Former national coordinator for AI.

10h10 – 11.15 – PANEL DISCUSSION

Questions from attendees will be asked to all speakers and three expert panelists:

- Isabelle Falque-Pierrotin
  President of the Autorité Nationale des Jeux
  Former president of the CNIL, saw the explosion of AI and its challenges (2009-2020).
- Nesrine Kaaniche,
  Assistant Professor, Télécom SudParis.
- Moez Draief,
  Vice President Data Science & Engineering, Capgemini.

ABSTRACTS OF RESEARCH ILLUSTRATIONS

Title: Algorithmic Fairness in Credit Scoring Models

Presenter: Christophe Pérignon (HEC Paris)

The development of AI stirred a passionate debate about the potential discrimination biases of the underpinning algorithms. Indeed, when assessing automatically the creditworthiness of borrowers, credit scoring models can place unprivileged groups, based for instance on their gender, race, citizenship or religion, at a systematic disadvantage. For instance, AI could lead to a significantly lower acceptance rate or higher interest rate for Afro-American loan applicants with a given annual income than for white loan applicants with the same level of income. Similarly, the Apple Pay app was publicly criticized for setting credit limits for female users at a much lower level than for otherwise comparable male users. While being sometimes illegal, such situations are often perceived as unethical and detrimental for the reputation of indicted companies. As a result, making sure that AI algorithms are fair is a top priority for governments and regulators, as demonstrated by recent regulation and white papers.

In this project, we develop a simple testing procedure for fairness metrics. Using a traffic-light approach, we qualify any scoring algorithm as either green (cannot reject the null of algorithm fairness), orange (can reject the null of algorithm fairness at a moderate confidence level), or red (can reject the null of algorithm fairness at a high confidence level). We then present an explainability technique, called Fairness Partial Dependence Plot, to identify the source(s) of the lack of fairness and mitigate fairness concerns. We illustrate the efficiency of our framework using a dataset of consumer loans and a series of machine-learning algorithms. While the focus of this paper is on credit scoring, our methodology can also be used in many other contexts in which AI algorithms are used to help making decisions.

Title: On Selection Bias and Fairness Issues in Machine-Learning

Presenter: Stephan Clémençon (Télécom Paris)

With the deluge of digitized information in the Big Data era, massive datasets are becoming increasingly available for learning predictive models. However, in many situations, the poor control of the data acquisition processes may jeopardize the outputs of machine-learning algorithms and selection bias issues are now the subject of much attention. Recently, the accuracy of facial recognition algorithms for biometrics applications has been fiercely discussed for instance, its monitoring over time revealing sometimes a predictive performance very far from what was expected at the end of the training stage. The use of machine-learning methods for designing medical diagnosis/prognosis support tools is currently triggering the same type of fear. Making the enthusiasm and the confidence for what can be accomplished by machine learning durable requires to revisit practice and theory both at the same time. It is precisely the purpose of this talk to explain and illustrate through real examples how to extend Empirical Risk Minimization, the main paradigm of statistical learning, when the training observations are biased, i.e. are drawn from distributions that may significantly differ from that of the data in the test/prediction stage. As expected, there is ‘no free lunch’: practical, theoretically grounded, solutions do exist in a variety of contexts (e.g. training examples composed of censored/truncated/survey data) but their implementation crucially depends on the availability of relevant auxiliary information about the data acquisition process. One should also have in mind that the ‘bias’ in machine-learning, as perceived by the general public, also refers to situations where the predictive error exhibits a huge disparity, to cases where the predictive algorithms are much less accurate for certain population segments than for others. If certain facial recognition algorithms make more mistakes for certain ethnic groups for instance, representativeness issues concerning the training data should not be incriminated solely: the variability in the error rates can be due just as much to the intrinsic difficulty of certain recognition problems or to the limitations of the state-of-the-art machine-learning technologies. As will be discussed in this talk, trade-offs between fairness and predictive accuracy then become unavoidable.

Title: Consumer Privacy Preservation – Firm’s and Regulator’s Perspectives.

Presenter: Ruslan Momot (Assistant Professor of Operations Management, HEC Paris)

The data privacy scandals, and breaches of recent years have exposed numerous risks resulting from the modern-day ubiquitous availability and accessibility of individual-level data. Thus are customers’ data routinely leaked and/or misused by adversaries against customers’ own interests. Even when anonymized, these data are not safe and can be “re-identified” to reveal a customer’s sensitive information. More advanced adversaries need not even gain direct access to the data; they can reconstruct sensitive individual-level information by observing decisions (regarding, e.g., prices and assortments) of the firms with which customers interacted.

We exploit the complexity of such data-dependent environments and build stylized mathematical models to explore the incentives of digital businesses to collect and protect users’ information and characterize how the revenue models of the businesses shape their equilibrium data policies. We find that, in general, companies collect more data and protect it less than what their customers would have wanted. Such misalignment of incentives is more pronounced for the companies with more data-driven revenue models. We show that a regulator can restore efficiency by introducing a two-pronged regulatory policy, which combines a minimal data protection requirement with a tax proportional to the amount of data collected. As another measure to preserve consumer privacy, we also develop new personalized pricing and assortment optimization privacy-preserving algorithms which can be used by companies engaged in revenue management. We show that when using our algorithms, privacy preservation can be achieved by the companies almost for free when sufficiently large historical data are available to them.

Title: Privacy and fairness concerns in the era of machine learning

Presenter: Catuscia Palamidessi (Ecole Polytechnique / INRIA)

The big data and machine learning technologies provide enormous benefits in many domains, ranging from economy and quality of service to scientific research. However, on one hand the collection and manipulation of personal data raises alarming privacy issues, on the other hand, the learning algorithms, especially the most powerful ones, result in decision-making devices that are often not transparent and risk to be unfair.

In this talk we illustrate one of the most popular and effective frameworks for privacy protection: differential privacy (DP). Besides the classic central and local DP models, we present our proposal for what we call the hybrid model. Namely, a mechanism to implement DP in the distributed case, i.e. in a scenario in which the data are distributed across different organizations, which do not wish to disclose their data, but still benefit from the advantages of combining their resources and enable federated learning.

Furthermore, we discuss some preliminary results about the relation between privacy and fairness in the context of machine learning. We show that, if on one hand privacy can go along and even enhance group fairness, on the other hand there is a tension between privacy and individual fairness, which may lead to a trade-off between these two ethical issues.