Big Data & Artificial Intelligence
Workshop 21 March - Summary

Dominique Guégan*
Université Paris 1 Panthéon-Sorbonne
Labex Refi

Alexis Bogroff†
Université Paris 1 Panthéon-Sorbonne

This document is a summary of the discussions which have been done, in Paris by the French team involved in the H2020 Fintech Risk Management project in association with the ACPR regulator, around the question of Big Data and Artificial Intelligence which is one of the key objective of this H2020 project. It is based on the 17 presentations done this day. The objectives were to identify the risks associated to the use of Big Data, the algorithms developed in that environment and the expectations of the regulators. First, we have defined the concepts under studies, then identify some risks for the both sides (Big Data and algorithms), with the comments provided during the workshop. Finally some case studies have been proposed by the banks, the fintechs and the researchers. The question of actual regulation is evocated with a lot of open questions.

The panel of persons which have been involved in the discussions is composed by the speakers whose list is at the end of the document and also the participants of this workshop. The workshop has been opened by Bertrand Peyret who is the Adjoint-General Secretary of ACPR which has welcome the workshop and Christophe Hénot who is the head of the project for France.

1 The concepts of Big Data and AI

1.1 Big Data

History from statistics to big data comes down into three parts: the individual, the population, and a compositional eras. The XVIIth century was the emergence of registries with Graunt' listing health statuses during the plague epidemic. The focus was on the fate of each individual. A 1760 Bernoulli and D'Alembert's debate on the innoculation of the smallpox ended the first period as it shifted the focus to the population's life expectancy. It was the beginning of the use of statistics to study society. Quételet and Galton later computed averages and standard deviations respectively (XIXth). Informations were written down at that time, and the arrival of typewriters and computers enhanced data handling by enabling manual inputs on spreadsheets (thousands of data points in 1990) and later automatic data flows on cloud

*dguegan@univ-paris1.fr
†alexis.bogroff@univ-paris1.fr

servers (terra and peta bytes stored in 2000). By combining statistics on the population with personal information, it enables a more precise profiling of each individual thanks to the recent datafication, digitalization and quantification. As we currently have the possibility to gather one million data points per day per individual, the profiling approach may be favored over the study of big trends.

What are the uses of such amounts of data? How to exchange and store it? What are the risks and what can regulators do? These questions are partly adressed in the following paper.

1.2 AI and algorithms

Artificial Intelligence (AI) was introduced by Alan Turing's research team in a 1950 paper *"Computing Machinery and Intelligence"*. AI is generally considered the big category encapsulating machine learning algorithms, from which supervised, unsupervised and reinforcement learning can be viewed as the main approaches as represented in figure 1.

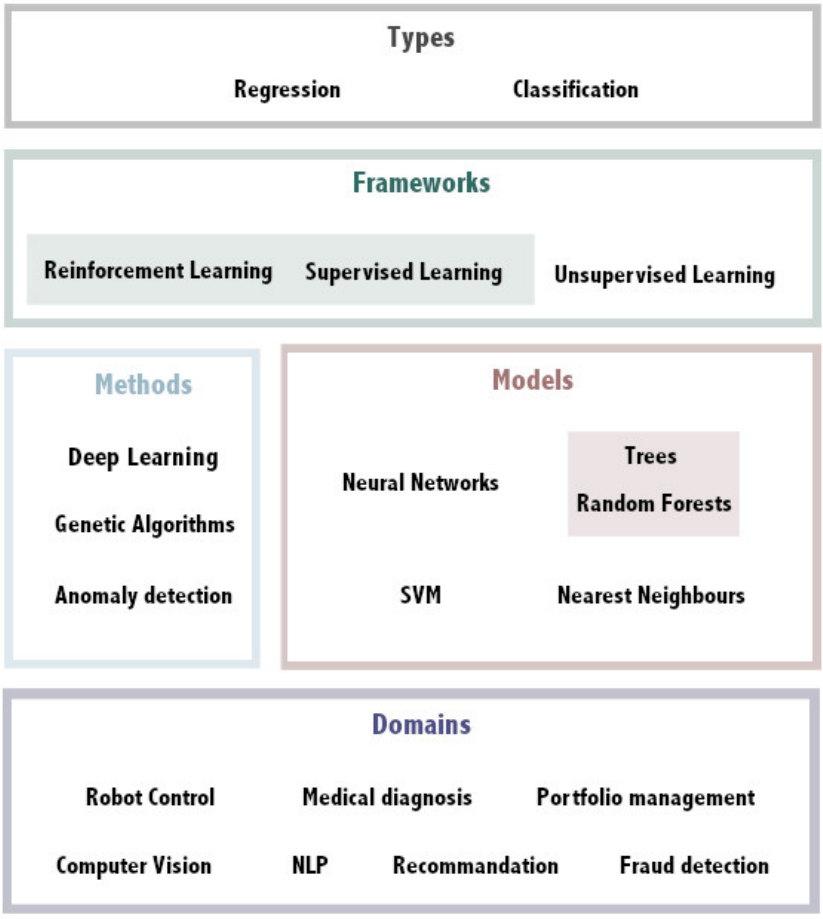

Figure 1: Machine Learning Categories

The fast development of machine learning algorithms and frameworks in the last decade creates both opportunities and risks. It not only concerns technical aspects of algorithms but also

includes their uses and potential biases and discriminatory drifts. How could cyber-security, bias and transparency be tackled, and what could be the role of supervisors to handle this double-sword challenge? Some risks encountered are described in the following section.

2 Risks identified on these subjects

A study of the risks' panel is important to better assess the need of regulations, control, and to find coherent measures that could support the development of these technologies without hindering them. In this section we attempt to present some risks with regard to data and algorithms discussed during the workshop.

2.1 Risks and Big Data

The analysis of the data appears as the main important task when we work with machine learning algorithms. Indeed, they will be determinant in the results of the algorithms and the decisions taken. It is considered that 80% of the job working with Big Data and algorithms is used to analyse and prepare the data.

2.1.1 Storage

The storage of data creates several technical challenges as it both requires efficiency, flexibility and security. Each can be partly addressed by third party operators that store data on their "cloud servers" (or platforms) and propose a diverse offer that could fit to many kind of uses. However the distant system increases the number of entry points for cyber-criminals that could stole information or modify data. Although this issue might be reduced by diminishing the number of intermediaries, it remains an area of constant research.

This raises many issues: who owns the data? Is it possible to get the data back, for example when the platform disappears, what happens? Regulation should clarify data holders' duties in order to answer to several questions regarding their storage politics, ethics, analysis of the data, supply chain in real time, data quality (automate process to detect errors - which could propagate), transformations, access rights, governance, and audit.

The current evolution leaves a few major economic actors (providing platforms), augmenting the risk of unbalanced commercial relations that could lead to artificially increased prices or limited services. It might also be a lever for the countries hosting such companies.

2.1.2 Control the evolution of sources

As soon as the data sets are stored on a platform (when the industries use this intermediary), very often algorithms are trained on an ongoing basis, thus the control of the data used is fundamental: what are the sources? Are the data modified (in particular their format)? When the data have been gathered, what are the main changes: new formula, data reprocessing added by the provider? What is the coherence of the data since the underlying process could evolve on the long run (communities moving from a social media to another, uses changing with new laws and habits)?

Is regulation possible?

2.1.3 Property

Property of data is an ethical and political issue rising in the consciousness of the population. It is mostly tackled by laws that were written along the expansion of the internet, but that do not expressively deal with machine learning or big data.

The recent General Data Protection Regulation (GDPR) in Europe has been implemented in 2016 for such purposes, but as the writing of this paper the regulators are attempting to better grasp these technologies in order to support and control them while preventing excessive restrictions that could hurt their development. Nevertheless some information is already forbidden for training models like ethnic group and residence address, but some correlated factors enable the reconstruction of the former like the locations of the shops and kind of restaurants that the person goes in. Breaking current laws can lead to penal sanctions. Educating on the subject and the presence of lawyers in companies involved in businesses of the data could help prevent unintended consequences.

Another point is how to deal with the right to oblivion? It is only partially tackled with the right to obtain the deletion of (some) personal data, but some indirect ways to identify a person are still under discussion/not punished. Some solutions could be to refuse to be profiled by such advanced methods, but how could it be possible for insurance companies for instance that could interpret it as a bad signal? Maybe forbid to treat it as a bad signal? How to verify that it is correctly applied? Is some regulation possible on this point?

2.1.4 Audit

In terms of regulation, the auditing of the company requires that any raw data, transformation on the data, algorithm used or results obtained be stored along the process. Since data used at different moments of the training may have different impacts on the behavior of an algorithm, a times-tamp history should be maintained for each data point. It would similarly concern each algorithm structure and weights that would be necessary to reconstruct the whole story of the data processing.

How to deal with such a huge amount of data? Is it realistic? Would it be useful? How to use such data to audit? Furthermore, how to address the random initialisation, sub-sampling and bagging issues? Otherwise, could it be possible to create specific methods for auditing machine learning algorithms (methods that do not require such precise data and would mostly be based on the current structure of an algorithm and its results)? The demand or the expectations of the regulator relatively to this question need to be specified.

2.1.5 Processing

As it has been pointed by several persons, data pre-processing represents 80% of a data scientist work. It has a critical influence on the results obtained by the algorithms, thus the quality of the data is crucial. It consists of several repetitive tasks like detecting and deleting outliers. Making such decisions is a highly sensible task as outliers could hinder the learning of the algorithm, but removing information also bias the data and eventually preclude the algorithm to aknowledge extreme cases. Normalizing data could in one side improve the learning efficiency and in another bias the information as it makes assumptions on the data distribution.

As further addressed in the audit section, each transformation of the data should be logged. Any missing information could be penalized during auditions. As it represents an outsized amount of records it should eventually be automatized or the regulator could adapt to request principal actions only. May a Github-like system help in keeping track of such transformations?

2.1.6 Bias and discrimination

Data could be biased with regard to a given objective. In a general sense, biases can eventually be positive (like a useful transfered knowledge), but the term is predominantly used to incriminate beliefs leading to a wrong understanding of a problem, or that are discriminating. In some cases biases could be intentionally introduced to treat political issues, such as choosing between equality and fairness, as both could be desired but are anti-ethical. The discussion around biased data and discriminating data is an opened route. Bias can be associated to some statistical property, discriminance is more associated to political choices and ethics. Bias can be introduced for instance by artificial intervention on the data set as explained further.

As soon as a database is created, the first application, concerning the persons (regarding consumption, crimes, recidives) could be the creation of profiles or segmentation for persons which can be discriminated. In that case it is the way by which the actors use the data (and then the algorithms which are defined with specific objectives) that creates the discrimination, the data is not discrimating by itself. As described through the evolution from statistics to big data (in section 1.1) the profiling approach over the study of big trends may now be favored. As soon as the profiles are entered and monitored by an algorithm, some ethic questions then arise. This last point has to be discussed from a political approach, and by the regulators.

2.1.7 Manipulation

Cyber-attacks or specific interventions on data could impede, mislead or destroy algorithms performances. Most notable attacks introduce falcified or misleading data. More advanced methods from financial markets like those using high-frequency trading (HFT) could directly impact the sources of data to manipulate the competitors' algorithms. Some hyper-trained Generative Adversarial Networks (GAN) could help both create, manipulate or detect such misleading data. The way in doing it is not so clear.

2.1.8 Choice of factors

The choice of factors is not a recent issue, but the current availability of large database at low-cost, combined with more advanced algorithms able to treat a more diverse set of data types, encourages the use of any available data to feed the models. However, models are mostly trained to find dependence rather than causalities. Hence the increased risk of spurious regressions or overfitting is real. It could partly explain the gap between Proof of Concepts (PoC) and operating systems. Some careful advices from econometrics would recommend using the least data sources.

Choosing factors is also at the crossroads of ethics and regulations. As stated above some sources are already forbidden since considered as providing personal information or discriminatory by nature. Probably this need to be more specified from regulators.

2.1.9 Regulation risk

How can a company be sure that it is allowed to use a given data source? Practitioners need to know the laws on the property of the data (RGPD, CNIL, etc.), but as the notion of Big Data is not defined in the law there is no right on data itself. Thus, a sui generis right exists to partly fill this gap, regulating data extraction from databases as detailed in section 4. However, nothing is said on the resulting owner of the patterns obtained after data processing: the producer of the database or the company applying its algorithms?

2.2 Algorithms

The risks which have been described previously relative to the data are very important and are not sufficiently documented for the users who often do not take care of all the previous points. Information and procedures could be involved to the potential users of Big Data sets. Thus, now considering the remarks previously done on the data sets, we analyse some specific risks which arrive with the use of algorithms in that context, indicating some possible solutions when they begin to be documented, if not proposing some routes for research.

2.2.1 Manipulation of algorithms: the role of the weights

Algorithms can be manipulated insidiously by amending their weights' value. Since Deep models have thousands of weights, few modifications could be imperceptible but have an important impact on outputs produced. As it could modify performances or orient results toward a specific objective, it would potentially wreck the whole (costly) learning of an algorithm.

Solution could be to implement means of control and restoration. In addition to secure connexions to the algorithm, independent model could verify that weights are always amended coherently (the process needs to be explicit).

2.2.2 Protection, patenting

As algorithms are considered "self-learning softwares" they can only be legally protected by the standard regulation on softwares. More complex issues are related to the property rights on patterns, since they are the result of both the algorithm (owned by the software company) and the data (owned by the database producer). This uncertainty could require the mentioning of rights on patterns in the contract established between the different parties.

Insurance could also protect related events such as the reputation risk in case of accidents with the algorithm. However, who is responsible for unintended consequences? Is it the software company (entity), its director or the developer? Also, who is responsible in case of an accident following a hardware failure? Would the software company be blamed for failing to predict its own breakdown, or for lacking a security system that should handle these situations? Is it possible to measure this kind of risks? What could be a good framework?

2.2.3 Transparency, interpretability

Algorithms transparency and interpretability is a major area of research as the recent development of deep methods provokes an explosion of the number of parameters and consequently

creates black boxes. Such methods need to be more interpretable to be used for important decisions, as choices and errors must be understood at least a posteriori (e.g. for court decisions, car driving and medical diagnoses).

Excessive transparency with outsiders can lead to business leaks, and a good balance is thus required to protect company's secrets of manufacture. This however has to be distinguished with the transparency of the model for internal control, which only has a positive effect.

Although unfair treatments could be detected by analysing the output of a model, a greater transparency could help preventing such issues upstream and better control models' biases. The research around these two subject - interpretablity and transparenty - has recently increased and some operation process could be developed to answer to these questions which are crucial for praticians, but also for the regulator in its relationship with the banking industry.

2.2.4 Operational risks: detecting algorithms' failures - Model's risk

The algorithms' performances and errors are almost necessarily handled during the training phase. However, depending on the type of task and availability of data, estimating the risks of failure of an algorithm can be challenging. In case of small or unbalanced datasets semi-supervised learning or SMOTE-like methods can help improve the generalization abilities, but for strongly unbalanced problems such as credit-scoring that are often too specific for gathering data from similar problems, the predictability rates of these methods are not yet compelling.

Failure detection could eventually be enhanced by a greater transparency of the models, or by providing more attention to the issue, such as using algorithms to predict other models' failures. These models could potentially be better at detecting failures than the algorithm itself if they transfered knowledge accross the different tasks to monitor. How to measure this kind of risk?

The failure of an algorithm is directly linked to the model's risk. The choice of an algorithm strongly determines outputs. In practice some models appear more efficient on certain types of tasks. Also some models are more transparent by design than others. The choices of the algorithm and its objective function are based on assumptions like the choice of the factors, but also on the criteria used to analyse the results provided by the different algorithms.

Also, if the model is handled by a platform, risks might arise from the following choices: the use of a black box model, forced calibrage, estimation, targets, parameters, cross-validation method and optimization techniques.

2.2.5 Human-machine interaction risk

Humans interacting with algorithms are not necessarily experts in the machine learning domain, thus its seems that a majority of models used in practice are solely relatively basic. Machines in the current economic state seem mostly used to augment humans' capabilities rather than for replacing them. Risks mostly come from the formation of the persons in charge of the management of these algorithms and the persons designed to take decisions. Thus, a major recommandation concerning the use of algorithms to take decisions whatever the field at which the research or enterprise want to apply them would be the obligation of the persons using the algorithm to take decisions, to understand how the algorithm works and how it does provide the results.

2.2.6 Systemic risks

A systemic risk could emerge with the rise of the use of similar algorithms. Since models are largely distributed and open-sourced we could observe the adoption of the same state-of-the-art model in equivalent businesses. Decisions made on the same object could modify the behavior of the underlying and potentially generate systemic risks on related environments, be it economic or car traffic domains. This kind of situations begins to be problematic, due to the up-to-date attraction of these subjects in the bank and also by the lack of formation of different actors in the bank.

Such a development would require risk measures of extremely dependent events, tools like copulas or dynamic networks could be useful.

2.2.7 Predicting extreme risks

Could algorithms detect black swans with bleu eyes (meaning events which never arrive)?

As models can be trained to understand the distribution of data, they could be used to at least measure and control skewness and kurtosis. Also, stress-GAN-tests could be implemented to strain operating models on plausible extreme scenarios, the philosophy can follows what has be done in Bale III with the introduction of stress tests to verify the safety of the bank in case of important shocks .

2.2.8 Mischievous uses

A great risk of fake information emerges with the rise of GANs and their ability to create fake data produced using a very similar distribution than real data. Faces and voices can be replaced on videos in an almost imperceptible way.

Algorithms could also be used to enhance cyber-attacks, more able to massively bypass standard protections.

Associating risks and their measures to this kind of situations is challenging.

3 Applications and case studies

During the workshop, researchers, practitioners and bankers presented some use cases of Big Data and machine learning algorithms that we classify into two categories: the applications improving tasks to streamline a company's operations, and those tackling more sensible business tasks. Most of the risks previously identified concern the latter class and strongly require more research, and perhaps regulations for a well defined framework (guidelines, risk measures, etc.).

Applications can either be implemented internally (as in some banks) or via cloud platforms. Cloud providers usually offer model performance and robustness indicators, up-to-date frameworks and models, traceability and reproductibility of scores. These features are essential to comply with audit requirements and need to be verified.

1. Applications improving the customer relationship:

- Answering general questions, helping or guiding the customer, making recommandations and suggesting personalized offers using Chatbots. In some applications, chatbots fully handle half the conversations.
- Improving services by analysising chatbots discussions through the detection of customer's negative comments, using sentiment analysis methods.
- Assisting in automatic e-mailing and redirection towards the right services using text analysis on mails.
- Gaining a better understanding of customers' thoughts and wishes from social media posts and conversations using text analysis methods (which may run continuously).

2. Applications improving the maintenance of the risk management system
   - Preventing breakdowns: a company's datasets can be used to improve the detection of operational risks since they are directly impacted. It is done using anomaly detection algorithms. The classical way to measure operational risks is still valuable.
   - Detecting frauds: organized fraud involve unusual customers connections and transaction frauds may create anomalies that can both be detectable on textual contents (using NPL, ML, Kohonen maps and DBScan), and some falsified documents (identity card with photoshop) can be blocked using Computer Vision.

     Models can train continuously to be competitive since fraud evolves rapidly.

     An issue with systems using auto-encoders for transaction fraud detection concerns the enterprise' level of knowledge: has it a perfect information? is all the information totally available to treat the problem?

3. Macroeconomic indicators prediction: using algorithms to modelize complex dependencies in a diverse set of factors. The sources could be: satellite to recognize transporters (boats), fields (type, quality, dryness), petroleum tanks (emptiness), military complexes and hospitals, social media (political elections), price variations (stocks, commodities, etc.), hotel prices and rate of corporate hotels (for prediction of local growth). From a regulatory approach, the origin of the data has to be known. Nevertheless the collect of the data cannot be regulated. In these cases it is necessary to process data of the highest quality and diversity.

4. Applications sensitive to ethical issues (see 2.1.6 for risks of bias and discrimination)
   - Crime recidive prediction (or other kinds of recidive): the data sources are: age, nature of the infraction (ignore ethnic group, etc.). It does not exist any consensus on the idea of using these programs: some are against (by principle) the idea of belonging to a machine for condamning people, other point out the fact that the algorithm used is not known. Even though it would be available, the common issue of algorithm's lack of transparency might remain unsolved. Two programs are currently in use in the USA: PSA (Public Safety Assessment) and COMPASS. How to deal with these issues? Enhancing transparency of models parly belongs to research. Refusing to be profiled by such advanced methods could be a solution, but is it possible for insurances which for instance would treat it as a bad signal?
   - Court decisions prediction: some studies show that results obtained could be biased (the question lie on the data sources). It could be necessary to open the case decisions and details making public the composition of the court, in order to verify that the

individual liberties are respected (can a person refuse this kind of procedure?) and to know how to handle extreme cases. It has been pointed that the use of multiple models on the same case (ensemble methods) will be necessary, in order to re-balance data and prevent some biased results.

- Default risk prediction: computation of default risks in order to provide loans to an enterprise. The conclusions are (i) to be relatively careful to the choice of the risks factors and to use a large dataset if machine learning modellings are considered, (ii) to use several algorithms competitively (it does not exist a unique modelling even for machine learning approaches), (iii) to use several criteria to take the final decision. When we use a machine learning system, it is also important to verify how it has been trained.

4 Regulation

We summarize some points raised during the workshop and provide a recent global information published April 8, 2019 by the European Commission (english[1], french[2]) around all the subjects discussed previously.

- Regulation is a central point as being both a source of risks for companies and a way to control risks for supervisors. Ledieu's blog[3] describes extensively these issues in a friendly way and we here focus on some fundamental notions of data property.

  Firstly, *"Big Data"* is not defined in European law texts. Thus, the regulation on data extraction mostly comes from a sui generis right. Rights on a database are owned by the *"producer"* of a database. This physical or moral person is considered the producer of the database if it is able to prove its investments regarding the collection of the data and setting up of the database. It then has a substancial, repeated and systematic right of extraction on its own database. The same terminology is thus used for proving illegal extractions from users. The producer can also grant full or partial extraction rights to its co-contractors or any other party.

  Thus, the user can extract data as long as it is not a substancial, repeated nor a systematic extraction. Alternatively, scraping is considered a criminal activity punishable by three years in prison and a 300,000 euros fine. Furthermore, as it corresponds (like any criminal sanction) to a civil fault, it thus have no limitation on responsibility.

  The second main right is the General Data Protection Regulation (GDPR) - EU2016/679 which defines rights on processed data. In these texts a personal data is defined as any directly or indirectly identifying data. Therefore, the use of the name and the living location but also IP, MAC, IMEI addresses that enable the identification of a terminal, and thus the owner of this terminal are prohibited.

  Finally, protecting machine-learning algorithms relates to software protection as it is considered an "autonoumously-learning" software (written in 1991 and re-written in the 2009 European directive 2009/24), and does not rely on the patenting of innovations.

[1] https://ec.europa.eu/commission/news/artificial-intelligence-2019-apr-08_en
[2] https://ec.europa.eu/commission/news/artificial-intelligence-2019-apr-08_fr
[3] https://www.ledieu-avocats.fr/le-droit-du-big-data

Some questions remain unanswered: who owns the resulting patterns, the database producer or the owner of the algorithm?

- On April 8, 2019 the European Commission unveiled a list of principles to follow to create "trustworthy" artificial intelligences. These recommendations are primarily aimed at protecting the most vulnerable groups such as children and persons with disabilities. They are also intended to protect confidentiality. On the site of the European Commission we find these informations:

  "Building on the work of the group of independent experts appointed in June 2018, the Commission launched a pilot phase to ensure that the ethical guidelines for Artificial Intelligence (AI) development and use can be implemented in practice. The Commission invites industry, research institutes and public authorities to test the detailed assessment list drafted by the High-Level Expert Group, which complements the guidelines. Today's plans are a deliverable under the AI strategy of April 2018, which aims at increasing public and private investments to at least €20 billion annually over the next decade, making more data available, fostering talent and ensuring trust. Artificial Intelligence can benefit a wide-range of sectors, such as healthcare, energy consumption, cars safety, farming, climate change and financial risk management. Artificial Intelligence can also help to detect fraud and cybersecurity threats, and enables law enforcement authorities to fight crime more efficiently. However, Artificial Intelligence also brings new challenges for the future of work, and raises legal and ethical questions.

  The essentials for achieving trustworthy Artificial Intelligence are (but no solution is provided):

  1. Trustworthy Artificial Intelligence should respect all applicable laws and regulations, as well as a series of requirements; specific assessment lists aim to help verify the application of each of the key requirements:
  2. Human agency and oversight: Artificial Intelligence systems should enable equitable societies by supporting human agency and fundamental rights, and not decrease, limit or misguide human autonomy.
  3. Robustness and safety: Trustworthy Artificial Intelligence requires algorithms to be secure, reliable and robust enough to deal with errors or inconsistencies during all life cycle phases of Artificial Intelligence systems.
  4. Privacy and data governance: Citizens should have full control over their own data, while data concerning them will not be used to harm or discriminate against them.
  5. Transparency: The traceability of Artificial Intelligence systems should be ensured.
  6. Diversity, non-discrimination and fairness: Artificial Intelligence systems should consider the whole range of human abilities, skills and requirements, and ensure accessibility.
  7. Societal and environmental well-being: Artificial Intelligence systems should be used to enhance positive social change and enhance sustainability and ecological responsibility.

8. Accountability: Mechanisms should be put in place to ensure responsibility and accountability for Artificial Intelligence systems and their outcomes."

Speakers of the Conference

Abiteboul Jeremie, Chief Product Officer, DreamQuark
Abraham Louis, Ecole Polytechnique, ETH Zurich, Qwant Care
Addo Peter, Agence Française de Développement (AFD)
Barry Laurence, Datastorm , Chaire Pari
Bounie David, Telecom ParisTech
Fliche olivier, Directeur Pole Fintech Innovation, ACPR
G'sell Florence, Université de Lorraine, IHEJ
Guégan Dominique, University Paris1 Panthéon - Sorbonne, LabEx ReFi
Hagen David, Commission de Surveillance du Secteur Financier (Luxembourg)
Hénot Christophe, University Paris1 Panthéon - Sorbonne, LabEx ReFi
Huynh Thanh-Long, QuantCube Technology
Ledieu Marc-Antoine, Avocats Associés, Constellation Avocats
Mostefa Djamel, Head of AI Orange Bank
Oukaci Farid, Banque de France
Peyret Bertrand, Secrétaire Général Adjoint de l'ACPR
Tillay-Doledec Bertrand, Head of Product, Scaled Risk
Yang Su, Pôle FinTech Innovation, ACPR