Welcome to our Hi! PARIS Fellows

Fellowships Projects
Presentation and discussion

September 27, 2021 • 9.00-10.30 AM

HI! PARIS is pleased to invite you to our welcome presentation and discussion with our Hi! PARIS Fellows on September 27, 2021.

Hi! PARIS is the new Center on Data Analytics and Artificial Intelligence for Science, Business and Society created by Institut Polytechnique de Paris (IP Paris) and HEC Paris and recently joigned by Inria Saclay.

Our 5 Hi! PARIS Fellows will be funded for their projects for the next 3 years, starting in September 2021. **Congratulations to Marie-Paule CANI (Ecole Polytechnique)** and **Arnak DALALYAN (ENSAE Paris)** as Fellowship Advanced, **Anna SIMONI (ENSAE Paris)** as Fellowship Consolidator, **Aymeric DIEULEVEUT (Ecole Polytechnique)** and **Xitong LI (HEC)** as Fellowship Starting!

This seminar is addressed to all faculties, engineers, and PhD students of HEC Paris, IP Paris, Inria Saclay and representatives of Hi! PARIS corporate donors, as well as to our scientific committee, governance committee and corporate advisory board members.

The event will be in a **hybrid format with speakers on site at Telecom Paris, Palaiseau (room 1A226)**. Please register through the Zoom webinar link below. If you want to join on site, send me an email to contact@hi-paris.fr !

Registration Link

Program

09:00-09:15 **Introduction** by Eric MOULINES and Nicolas VIEILLE, Hi! PARIS Scientific Co-Directors

09:15-10:15 **Presentation of the Fellows' projects**

- **09.15-09.30** **Aymeric DIEULEVEUT** - "Create new algorithms for federated learning, supported by theoretical guarantees"
- **09.30-09.45** **Xitong LI** - "The Impacts of Algorithmic Decision-makers, Advisors, and Recommenders on Business and Society"
- **09.45-10.00** **Arnak DALALYAN** - "Statistical Analysis of Generative Models: Sampling Guarantees and Robustness"
- **10.00-10.15** **Marie-Paule CANI** - "CREATIVE AI : CREating and Animating Testbeds of Ideas, through Virtual Environments with Artificial Intelligence"

10.15-10.30 **Open discussion**

Presentation's description

For more details on the Fellowships: https://www.hi-paris.fr/2021/06/21/fellowships-2021-results/

Fellowship Advanced

Marie-Paule CANI, Ecole Polytechnique

Hi! PARIS project: "CREATIVE AI: CREating and Animating Testbeds of Ideas, through Virtual Environments with Artificial Intelligence"

Research field: Computer Graphics: Smart 3D Graphics, Computer Animation, Interaction

Abstract: Scientists cannot understand or reason about the world without some, at least mental, visual representation. The latter not only heightens their intuition, but also helps them structure the observed phenomena into a series of simpler stages. Over centuries, scientists used static sketches for expressing their thoughts, despite of the difficulties for 2D drawings to convey animated 3D shapes. They had to iteratively modify these representations as their mental picture crystallized. With the advent of visual computing, one would expect digital images to have replaced such traditional practices. Unfortunately, while captured images cannot convey a scientist's vision in a meaningful way, the use of 3D modeling also fails to solve the problem: Indeed, designing animated 3D contents requires an iterative process between the scientist and a digital artist trained on dedicated tools, leading to weeks of tedious work and successive refinements. Moreover, the resulting 3D contents are sterile, they cannot be used as a testbed to play with ideas and improve them. Could digital images be turned into a creative media, more expressive but as simple to use as a pen, enabling scientists to directly express these visions in the form of interactive 3D environments, manipulate them, test different hypotheses through interaction and inject new data on the fly, to get inspiration and refine their thoughts? This is the goal of my project. Achieving it will make scientific thinking directly possible in digital form, from early intuitions to progressive refinement, testing and finalization of an idea. Recent advances in Artificial Intelligence open the way to such a revolution of contents creation. However, AI should be revisited from a user-centered perspective. This will imply bridging the gap between (i) Expressive gesture-based creation, based on sketching, sculpting or smart copy-paste, developed so far for isolated 3D shapes; (ii) Procedural models for visual simulation, conveying prior knowledge through rules and constraints, but often slow and difficult to control; and (iii) Learning processes that build on data and examples to quickly calibrate models or generate contents though analogies. In contrast with intelligent systems that would tackle creation for us, the goal of the resulting "Creative AI" is to support and enhance users' creativity. Scientifically, achieving Creative AI calls for a new convergence between Artificial Intelligence, Graphical Simulation in 3D, and Human Computer Interfaces. In terms of impact, we believe that the resulting extension of virtual environments into "creative testbeds" enabling to convey, explore, test, and refine mental visions, will become a key support for scientists willing to think about a problem, discuss and share their vision with colleagues, or communicate their knowledge with the public. Such new creative media may also serve the needs of engineers or designers willing to explore coarse to fine creation of virtual prototypes, educators aiming at quickly conveying some insight, or any people willing to give live, test and share imaginary 3D worlds. Therefore, this project opens new horizons for science, technology, and society.

Arnak Dalalyan, ENSAE Paris

Hi! PARIS project: "Statistical Analysis of Generative Models: Sampling Guarantees and Robustness"

Research field: Nonparametric and high-dimensional Statistics, Statistical Learning, Computer Vision

Abstract: Statistical methods are omnipresent in the most powerful artificial intelligence technologies. The algorithms that are currently used rely on high-dimensional and highly-overparameterized models (such as deep neural networks) for which classical theoretical results do not lead to a good understanding of the empirically observed phenomena. The goal of this project is to develop mathematical tools that are tailored to the high-dimensional and highly-parameterized models by focusing on model averaging, approximate sampling, generative modeling and robustness. While most recent approaches to theoretically justify the success of deep learning try to quantify the performance of the trained model as a solution to an optimization problem, we intend to study this question through the lens of model averaging and sampling from a given distribution. The project

also intends to investigate optimality properties in generative models such as conditional generative models and cycle-generative models. More precisely, we will focus on getting finite-sample statistical guarantees which will clearly highlight the impact of a suitably defined notion of intrinsic dimension (as opposed to the ambient dimension) on the risk. This kind of guarantees are paramount in machine learning to ensure the reliability of the obtained algorithms. Furthermore, we will pay special attention to the stability and robustness properties: robustness to the model mis-specification and robustness to the presence of outliers in data.

Fellowship Consolidator

Anna SIMONI, ENSAE Paris

Hi! PARIS project: "Macroeconomic nowcasting with high-dimensional Google Search data: theory and practice"

Research field: Econometrics, High Dimensional Econometrics, Bayesian Inference, Google search data, semi- and non-parametric econometrics, nowcasting

Abstract: This project will develop machine learning tools tailored for a time-series environment with the aim of nowcasting macroeconomic and financial aggregates by using Google Search data together with official data. Nowcasting aims at providing an evaluation of the macroeconomic and financial activity in real time, which is crucial for the policy-maker to promptly react to changes in the economy with appropriate counter-cyclical economic policies. Indeed, the problem that forecasters face is that official series are often published with a delay. We are interested in exploiting in the best possible way the economic informational content of Google Search data and in seeing how the combination of these data with the official series (when available) can improve prediction accuracy. Because Google Search data contain an ultra-high number of variables (categories) compared with the time dimension, our first goal is to shape pretesting procedures that retain only the more important Google Search variables to improve nowcasting accuracy.
We will then develop theoretical properties of procedures that apply different penalization and factor-based methods to the targeted pre-selected Google Search data. These properties will take into account different temporal dependency structure and correlation in the data.
A further objective will be to show that Google Search data can be useful to detect structural breaks and to propose a procedure to estimate the break points and the coefficients of the corresponding models. A final goal will consist in reducing the mean squared forecasting error of the nowcasts through bagging methods that we will develop for this particular framework involving nowcasting and time-series environment.
Overall, this project will contribute to the machine learning literature in econometrics and economics from both a theoretical and empirical point of view.

Fellowship Starting

Aymeric DIEULEVEUT, Ecole Polytechnique

Hi! PARIS project: "Create new algorithms for federated learning, supported by theoretical guarantees"

Research field: Nonparametric and high-dimensional Statistics, Statistical Learning, Computer Vision

Abstract: The design of new algorithms for artificial intelligence is one of the major challenges of our time. The extraordinary progress made in recent years has been enabled by a combination of factors: on the one hand, the simultaneous growth of available data sets and computing power; on the other hand, algorithmic innovations that have played a decisive role in making it possible to move to the next scale. During the last four years, a new domain of the field has gained a lot of importance, Federated Learning. In Federated Learning, several organizations or devices seek to collaboratively train a model under the orchestration of a central server, while keeping individual datasets on their respective local storage. Federated Learning has emerged as a fundamental setting to tackle new societal and industrial challenges in machine learning. Indeed, privacy has become a major concern for both society (individuals participating into the training want to protect their privacy) and industries

(facing legal constraints preventing them from exploiting valuable data). Overall, it becomes necessary to train the models without centralizing the data, either because of those privacy constraints, or sometimes because extremely large datasets have to be distributed over networks of storing devices. Consequently, new optimization challenges are arising, taking into account communication constraints, and new opportunities to improve the models to adapt to the users are being explored. Developing new algorithms for Federated Learning is thus a key challenge. It will simultaneously positively impact society, by protecting individual data and restoring public trust and confidence in machine learning technologies, and unlock countless novel opportunities of collaboration between entities willing to collaborate without centralizing their datasets; a crucial situation in medical applications, fraud detection, IOT, and many other domains. In this proposal, we describe four pivotal challenges of Federated Learning: communication constraints, statistical heterogeneity, missing data and privacy. Then, we identify four research directions: in each direction, we tackle a combination of several of those challenges. Our goal is to build the next generation of algorithms for large scale Federated Learning, supported by strong theoretical guarantees, and practical implementations together with open-source code. These new algorithms, methods and guarantees constitute a key step to build reliable Federated Learning architectures, and are expected to bring major advances to the field, by improving scalability and efficiency of the methods, and enabling fruitful collaborations that would not have occurred otherwise. To do so, we will rely on a precise mathematical understanding of optimization algorithms and of the statistical tradeoffs of large scale learning. The resources of the fellowship are crucial to complete the fellows' team and accelerate the exploration of these competitive topics, in order to submit an ERC proposal in two years.

Xitong LI, HEC Paris

Hi! PARIS project: "The Impacts of Algorithmic Decision-makers, Advisors, and Recommenders on Business and Society"

Research field: Economics of information technologies (social media, crowdfunding, digital marketing, online education), FinTech

Abstract: An algorithm is a series of operations commonly executed by computers to automatically perform data processing, predictions and reasoning. Given algorithms' abilities to process "big data" and perform the tasks consistently, algorithms are expected to outperform humans in many circumstances. We therefore see vast implementations of algorithmic decision-makers, advisors, and recommenders in our business and society, with the least mature category of algorithmic decision-makers and the most mature category of algorithmic recommenders. Therefore, this project proposal comprises three work packages (WPs), each of which examines the impact of one of the three categories of algorithm-based applications.

WP1 focuses on examining algorithmic decision-makers, specifically, algorithmic investment robots in an online peer-to-peer lending market. As machine learning technologies advance, algorithms increasingly make decisions on behalf of humans. Despite the considerable body of literature on social influence from humans to humans, to the best of my knowledge, no research examines the extent to which humans' decisions reflect the influence of prior decisions made by algorithms, referred to as "algorithmic influence". Therefore, the proposed research project P1 in WP1 aims to explore: *Algorithmic Influence: How Does Algorithmic Investment Influence Human Investment?*

WP2 focuses on examining algorithmic advisors which give algorithm-based judgment and opinions to assist humans in their judgment. A recent stream of experimental studies find that individuals have a tendency of algorithm appreciation in the sense that they exhibit higher degree of advice taking when the advice is from an algorithm than the identical advice from humans. Yet, the extant research also shows that the extent of algorithm appreciation depends on contingent factors, such as the indication of algorithm's prediction performance. The transparency of an algorithm (or algorithmic transparency) refers to the degree to which the information related to the algorithm is revealed and also how such information is presented. The information can be about the algorithms' prediction performance or about the input variables used by the algorithms. Therefore, this work package WP2 includes two proposed research projects: Algorithmic Appreciation: Does Presenting Prediction Performance Matter? ; How Does Algorithmic Transparency Influence Adoption of Algorithmic Advisors?

WP3 focuses on examining algorithmic recommenders, specifically, algorithmic product recommenders in online retailing. Despite the considerable body of literature on algorithmic product recommendations, this work package WP3 identifies three literature gaps and accordingly three proposed research projects: Retargeted versus Generic Product Recommendations: When is it Valuable to Present Retargeted Recommendations? ; How Do Recommender Systems Lead to Consumer Purchases? A Causal Mediation Analysis of a Field Experiment ; How Beneficial are Recommendations to Consumers? Estimates of Relative Benefits of Product Recommendations to Consumers and Retailers