

Hi! PARIS is pleased to propose its second summer school on **July 4-7, 2022**.

Hi! PARIS is the interdisciplinary center for **Data Analytics and Artificial Intelligence for Science, Business and Society**. Founded by HEC Paris and Institut Polytechnique de Paris (IP Paris) and joined in 2021 by Inria, the center sets a standard of excellence for high-level research projects, educational programs, and business applications.

The Hi! PARIS Summer School 2022 on AI & Data for Science, Business and Society covers a wide range of topics in Artificial Intelligence and Data Science from a variety of perspectives. **This summer school offers courses that range from introduction to deep reinforcement learning to Intelligent Risk Management, Image Recognition Using Deep Learning, Optimal Transport for Machine Learning and Supervised learning on multivariate brain signals**. Our summer school should be of interest to PhD track students and final year students going for a PhD, to current PhD students, to academics and to research engineers who want to expand their knowledge in these areas.

The Hi! PARIS Summer School will be held at **Telecom Paris, in Palaiseau**.

Monday, July 4

08:00 – 09:00 Welcome coffee

09:00 – 09:15 Inauguration

09:15 – 10:45 Industry Panel

Opportunities and Challenges with AI and Data Science

with **Hi! PARIS Corporate Donors' representatives**, introduced and moderated by **Jean-Eduard Colliard**, HEC Paris

10:45 – 11:00 Coffee Break

11:00 – 12:30 Tutorial*

Tutorial 1A _ Part 1

Data in Finance : FinTech Lending

Johan Hombert
HEC Paris
Chairperson: **Jean-Eduard Colliard**

Tutorial 1B _ Part 1

Optimal Transport for Machine Learning

Rémi Flamary
Ecole Polytechnique
Chairperson: **Arnak Dalalyan**

12:30 – 13:30 Lunch Break

13:30 – 15:00 Tutorial*

Tutorial 1A _ Part 2

Data in Finance : FinTech Lending

Johan Hombert

Tutorial 1B _ Part 2

Optimal Transport for Machine Learning

Rémi Flamary

15:00 – 15:30 Coffee Break

15:30 – 16:30 Keynote

Opinion dynamics in online social networks: models and computational methods

Aristides Gionis, WASP professor at KTH Royal Institute of Technology and an adjunct professor at Aalto University
Chairperson: **Mauro Sozio**

16:30 – 18:00 **Social Time**
Conference Networking coffee

* The tutorials are organized in two parallel tracks: Track A "Data Science for Business and Society" and Track B "Theory and methods of IA".

Tuesday, July 5

08:00 – 08:30 Welcome coffee

08:30 – 10:00 Tutorial

<p style="text-align: center;">Tutorial 2A _ Part 1</p> <p style="text-align: center;">Operationalizing AI Regulation</p> <p>David Restrepo HEC Paris Winston Maxwell Télécom Paris <i>Chairperson: David Restrepo-Amariles</i></p>	<p style="text-align: center;">Tutorial 2B _ Part 1</p> <p style="text-align: center;">Supervised learning on multivariate brain signals</p> <p>Alexandre Gramfort INRIA <i>Chairperson: Arnak Dalalyan</i></p>
--	---

10:00 – 10:30 Coffee Break

10:30 – 12:00 Tutorial

<p style="text-align: center;">Tutorial 2A _ Part 1</p> <p style="text-align: center;">Operationalizing AI Regulation</p> <p>David Restrepo Winston Maxwell</p>	<p style="text-align: center;">Tutorial 2B _ Part 2</p> <p style="text-align: center;">Supervised learning on multivariate brain signals</p> <p>Alexandre Gramfort</p>
---	--

12:00 – 13:00 Lunch Break

13:00 – 14:00 Keynote

<p>From Artificial Intelligence to Augmented Intelligence</p> <p>Balaji Padmanabhan University of South Florida, Anderson Professor of Global Management, director of the Center for Analytics & Creativity, a professor in the School of Information Systems and Management <i>Chairperson: Shirish C. Srivastava</i></p>
--

14:00 – 15:30 Tutorial

<p style="text-align: center;">Tutorial 3A _ Part 1</p> <p style="text-align: center;">Image Recognition Using Deep Learning : Implementation and Application</p> <p>Mitali Banerjee HEC Paris <i>Chairperson: Shirish C. Srivastava</i></p>	<p style="text-align: center;">Tutorial 3B _ Part 1</p> <p style="text-align: center;">Hybrid Artificial Intelligence and Image Understanding</p> <p>Isabelle Bloch Sorbonne University - LIP6 <i>Chairperson: Mauro Sozio</i></p>
--	--

15:30 – 16:00 Coffee Break

16:00 – 17:30 Tutorial

<p style="text-align: center;">Tutorial 3A _ Part 2</p> <p style="text-align: center;">Image Recognition Using Deep Learning : Implementation and Application</p> <p>Mitali Banerjee</p>	<p style="text-align: center;">Tutorial 3B _ Part 2</p> <p style="text-align: center;">Hybrid Artificial Intelligence and Image Understanding</p> <p>Isabelle Bloch</p>
--	---

17:30 – 18:30 Poster presentation

Wednesday, July 6

08:00 – 08:30 Welcome coffee

08:30 – 10:00 Tutorial

<p style="text-align: center;">Tutorial 4A _ Part 1</p> <p style="text-align: center;">Impact of Privacy Regulation on Online Advertising Market : GDPR in Europe</p> <p>Klaus Miller Hi! PARIS Chair holder at HEC Paris <i>Chairperson: David Restrepo-Amariles</i></p>	<p style="text-align: center;">Tutorial 4B _ Part 1</p> <p style="text-align: center;">An Introduction to Deep Reinforcement Learning</p> <p>Corentin Tallec DeepMind <i>Chairperson: Anna Korba</i></p>
---	--

10:00 – 10:30 Coffee Break

10:30 – 12:00 Tutorial

<p style="text-align: center;">Tutorial 4A _ Part 2</p> <p style="text-align: center;">Impact of Privacy Regulation on Online Advertising Market : GDPR in Europe</p> <p>Klaus Miller</p>	<p style="text-align: center;">Tutorial 4B _ Part 2</p> <p style="text-align: center;">An Introduction to Deep Reinforcement Learning</p> <p>Corentin Tallec</p>
---	--

12:00 – 13:00 Lunch Break

13:00 – 14:00 Keynote

Mathematics of Neural Networks in the Billion-parameter Age

Joan Bruna

Associate Professor at Courant Institute, New York University (NYU)

Chairperson: Arnak Dalalyan

14:00 – 15:30 Academic Round Table

AI and Society

With the Hi! PARIS Summer School 2022 keynote speakers, introduced and moderated by Anna Korba, IP PARIS – ENSAE Paris

15:30 – 16:00 Coffee Break

16:00 – 17:30 Student Program

Who Better To Give Professional Advice Than Students Who Have Gone Through The Same Situations As You?

18:00 – 23:45 Social Event

Thursday, July 7

08:00 – 08:30 Welcome coffee

08:30 – 10:00 Tutorial*

Tutorial 5A _ Part 1 Decision-making Under Uncertainty Julien Grand-Clément Hi! PARIS Chair at HEC Paris <i>Chairperson: Jean-Edouard Colliard</i>	Tutorial 5B _ Part 1 Deep Learning for Audio Processing Geoffroy Peeters IP PARIS – Telecom Paris <i>Chairperson: Jean-Arnak Dalalyan</i>
--	---

10:00 – 10:30 Coffee Break

10:30 – 12:00 Tutorial

Tutorial 5A _ Part 2 Decision-making Under Uncertainty Julien Grand-Clément	Tutorial 5B _ Part 2 Deep Learning for Audio Processing Geoffroy Peeters
---	--

12:00 – 13:00 Lunch Break

13:00 – 14:00 Keynote

Developing Artificial Intelligence for the Public Good Helen Margetts Professor of Society and the Internet at the University of Oxford <i>Chairperson: David Restrepo-Amariles</i>
--

14:00 – 15:30 Tutorial*

Tutorial 6A _ Part 1 Intelligence Risk Management: Graph-Based Anomaly Detection Using the MDL Principle Aluna Wang Hi! PARIS Chair at HEC Paris <i>Chairperson: Jean-Edouard Colliard</i>	Tutorial 6B _ Part 1 Reliable Decision Making and Causal Inference with Kernels Krikamol Muandet Max Planck Institute for Intelligence Systems <i>Chairperson: Anna Korba</i>
--	---

15:30 – 16:00 Coffee Break

16:00 – 17:30 Tutorial

Tutorial 6A _ Part 2 Intelligence Risk Management: Graph-Based Anomaly Detection Using the MDL Principle Aluna Wang	Tutorial 6B _ Part 2 Reliable Decision Making and Causal Inference with Kernels Krikamol Muandet
---	--

17:30 – 18:30 Poster Award

18:00 Closing Session

Wi-fi connection on Campus Telecom

Login telecom207043-1807
 Password Wq1BHTsh

Sessions Description

The sessions are described on the dedicated page of the website: <https://www.summerschool.hi-paris.fr/program/sessions-description/>

For more details on the speakers: <https://www.summerschool.hi-paris.fr/list-of-speakers/>

Keynotes

Keynote 1 – Opinion dynamics in online social networks: models and computational methods

Speaker: Aristides GIONIS

Abstract: Online social networks are widely used nowadays by people to engage in conversations about a variety of topics. Over time, these discussions can have a significant impact on people's opinions. In this talk we present an overview of models that have been proposed in the literature to capture how information spreads and how opinions form in online social media. One of our objectives is to obtain a better understanding of adverse social phenomena, such as increase in polarization and creation of filter bubbles. We then present some of the computational challenges that have arisen recently in this domain. In particular, we discuss mediation strategies for maximizing the diversity of the content of users via recommendations and prioritizing their feed in order to reduce polarization. Finally, we study the question of whether an adversary can sow disagreement in a social network by influencing the opinion of a small set of users.

Keynote 2 – From Artificial Intelligence to Augmented Intelligence

Speaker: Balaji PADMANABHAN

Abstract: This talk will first present an overview of Artificial Intelligence over the years leading to where we are today, and use that historical context to discuss some of the successes and failures we have seen over the years. With this backdrop, we will set the stage for thinking about designing for “augmented intelligence” not just artificial intelligence; where we want to think about tackling more complex business and societal problems (than image recognition, say) using a combination of data, algorithms and people. In addition to providing an overview of some important recent work in this context we will present a complex systems perspective into this issue, and show how such a perspective can be useful to design, develop, evaluate and refine newer augmented intelligence methods going forward.

Keynote 3 – Mathematics of Neural Networks in the Billion-parameter Age

Speaker: Joan BRUNA

Abstract: The pace of progress of large-scale machine learning keeps increasing, towards even bigger models and datasets, producing astonishing results along the way in data-heavy domains such as text and images. Such rapid progress also leaves our mathematical understanding further behind, to the extent that one wonders whether it will ever catch up? In this talk, we will raise salient questions about this trend while zooming-in on technical snippets, converting approximation properties of

transformers, mathematical aspects of score-based diffusion generative models, and optimization aspects of learning semi-parametric models.

Keynote 4 – Developing Artificial Intelligence for the Public Good

Speaker: Helen MARGETTS

Abstract: Most Artificial Intelligence is developed by and for the private sector. This talk will focus on what can happen when we think about AI from a public sector perspective. How can AI be used to improve policymaking, public services and governance? What are the 'wicked' public policy problems that AI might help to solve? Drawing on research underway at the Public policy program at The Alan Turing Institute for Data Science and AI in the UK, the talk will explain the tasks for which data science and AI are particularly suited. It will show how the use of these data-driven technologies can foster government innovation, optimize resource allocation, and highlight longstanding injustices in public decision-making. Developing and using AI in the public sector might help to make governments more efficient, effective, fair and resilient than ever before.

Tutorials

The tutorials are organized in two parallel tracks:

- Track A “Data Science for Business and Society”
- Track B “Theory and methods of IA”.

Track A “Data Science for Business and Society”

Tutorial 1A – Data in Finance: FinTech Lending

Speaker: Johan HOMBERT

Abstract: This tutorial includes a short lecture followed by an interactive game in which participants play the role of a FinTech lender. Context: Banks and insurers increasingly use alternative data and machine learning to screen consumers and price products. For example, a FinTech using digital footprints to predict default will have a competitive edge over traditional banks. However, there are important pitfalls to avoid when using alternative data and machine learning to score consumers, such as the winner’s curse, the risk of discrimination and the Lucas critique. This tutorial and its interactive game provide an introduction to these issues.

Tutorial 2A – Operationalizing AI Regulation

Speaker: David RESTREPO AMARILES & Winston MAXWELL

Abstract: How will Europe’s future AI regulation impact the design, testing and use of AI applications such as credit scoring, recruitment algorithms, anti-fraud algorithms and facial recognition? We will explore how AI concepts such as explainability, fairness, accuracy, robustness and human oversight will be implemented into the future regulation, and how the regulation compares to other international standards on trustworthy AI. The course will focus on two concrete use cases, facial

recognition and credit scoring, to see how the European regulatory framework would apply throughout the lifecycle of the project, walking students through the process of creating a risk management system, including the impact assessment on potential risks for safety and fundamental rights, developing a list of requirements, testing, performance parameters, documentation and human oversight mechanisms. We'll explore the potential friction between the European AI Act and other regulatory frameworks such as the European General Data Protection Regulation (GDPR) and lead a debate on how the future regulation will impact AI innovation and research in Europe.

Tutorial 3A – Image Recognition Using Deep Learning: Implementation and Application

Speaker: Mitali BANERJEE

Abstract: This 3-hour module will offer a hands-on introduction to deep learning based image recognition tools. Participants will gain familiarity with preparing and importing images into software (python) and applying one of the foundational deep learning architectures to classify the images and create vector representations. We will discuss different applications of the output of deep learning tools to extract managerial and scientific insights. In particular, the course will discuss applications of these tools to create large-scale measures that have otherwise proven to be elusive to measure or susceptible to bias in measurement.

Tutorial 4A – Impact of Privacy Regulation on Online Advertising Market: GDPR in Europe

Speaker: Klaus MILLER

Abstract: We will discuss the impact of privacy regulation on the online advertising market and specifically focus on the case of the European Union's General Data Protection Regulation (GDPR). Specially, participants of this tutorial will learn: (1) Why and how the European General Data Protection Regulation (GDPR) impacts the online advertising market, particularly advertisers, publishers and users. (2) How advertisers and publishers leverage users' personal data to pursue their goals. (3) Which aspects of the GDPR are most relevant for advertisers, publishers and users. (4) How complex it is to go through the process of obtaining user permission for personal data processing, and how IAB's Transparency and Consent Framework (TCF) intends to help. (5) How many firms a publisher provides with access to its user's data, and how long it takes a user to respond to all permission requests. (6) Which developments are taking place with regard to personal data processing, among players in the online advertising industry, as well as among regulators and consumer protection agencies. Anyone interested in learning how and why the online advertising industry benefits from using personal data, and how the GDPR impacts this practice should attend this tutorial.

The tutorial is based on the book “The Impact of the General Data Protection Regulation (GDPR) on the Online Advertising Market” available completely for free at www.gdpr-impact.com.

Tutorial 5A – Decision-making Under Uncertainty

Speaker: Julien GRAND-CLEMENT

Abstract: The goal of this tutorial is to understand how uncertainty impacts classical decision-making models and the operational and business consequences. Any decision model that is data-driven may face uncertainty due to errors in the data, in the modeling assumptions, or due to the inherent randomness of the decision process. Overlooking this uncertainty may lead to decisions that are suboptimal, unreliable, or, in some crucial applications, practically infeasible and dangerous for the users. In this tutorial, we will learn to (1) estimate the uncertainty given a decision problem and a dataset, and (2) mitigate the impact of uncertainty with a robust approach. As an application, a robust portfolio management problem will be investigated in detail, though we will see that the problem of uncertainty arises in many (if not most) real decision settings. This tutorial is structured as follows:

1. How to estimate the uncertainty in a decision model?

Motivating examples: what is the practical impact of uncertainty?

Wrong images classification, variability in demands for supply chains, artificial intelligence in healthcare, Tesla auto-driving, robotics, maintenance, inventory optimization, facility location, project management, etc.

Introduction of the running example: portfolio management.

Understanding the origin of the uncertainty: poor data, little data, is the uncertainty inherent to the application? When do we need to take it into account?

Risk-sensitive decisions vs. parameter uncertainty.

2. How to estimate the uncertainty? Examples with simulations with Colab and synthetic data for the portfolio management problem.

How to mitigate the impact of uncertainty in practice? Robust portfolio management.

Deterministic approach: pessimism in parameters estimations.

Robust and distributional robust approach: how to obtain decisions with guarantees of good performances.

Evidence from simulations with Colab: trade-offs nominal performances vs. worst-case performances for the portfolio management problem. How to deal with variability?

(Time-permitting) Two-stage decision-making: how to act when uncertainty is revealed over time?

Tutorial 6A – Intelligent Risk Management: Graph-Based Anomaly Detection Using the MDL Principle

Speaker: Aluna WANG

Abstract: Risk management encompasses the identification, analysis, and response to risk factors arising over the life of a business. Recognizing patterns and detecting anomalies in big data can be critical to effective risk management. While numerous technologies for spotting anomalies in collections of multi-dimensional data points have been developed in the past years, anomaly detection techniques for structured graph data have lately become a focus. Why do we need to use graph-based approaches to anomaly detection? What are some of the high-impact applications of graph-based anomaly detection in risk management? How can we develop and deploy graph-based anomaly detection techniques for financial transaction data? This short course answers the above questions by introducing two general, scalable, and explainable anomaly detection models, with a focus on the use of graphs and the minimum description length (MDL) principle. The course also discusses how to deploy these techniques and use them for risk management.

Track B “Theory and methods of IA”

Tutorial 1B – Optimal Transport for Machine Learning

Speaker: Rémi FLAMARY

Abstract: This tutorial aims at presenting the mathematical theory of optimal transport (OT) and providing a global view of the potential applications of this theory in machine learning, signal and image processing and biomedical data processing. The first part of the tutorial will present the theory of optimal transport and the optimization problems through the original formulation of Monge and the Kantorovitch formulation in the primal and dual. The algorithms used to solve these problems will be discussed and the problem will be illustrated on simple examples. We will also introduce the OT – based Wasserstein distance and the Wasserstein barycenter that are fundamental tools in data

processing of histograms. Finally, we will present recent developments in regularized OT that bring efficient solvers and more robust solutions.

The second part of the tutorial will present numerous recent applications of OT in the field of machine learning and signal processing and biomedical imaging. We will see how the mapping inherent to optimal transport can be used to perform domain adaptation and transfer learning. Finally, we will discuss the use of OT on empirical datasets with applications in generative adversarial networks, unsupervised learning, and processing of structured data such as graphs.

Tutorial 2B – Supervised learning on multivariate brain signals

Speaker: Alexandre GRAMFORT

Abstract: Understanding how the brain works in healthy and pathological conditions is considered as one of the major challenges for the 21st century. After the first electroencephalography (EEG) measurements in 1929, the 90’s was the birth of modern functional brain imaging with the first functional MRI (fMRI) and full head magnetoencephalography (MEG) system. Presently new tech companies are developing new consumer grade devices for at home recordings of neural activity. By offering noninvasively unique insights into the living brain, these technologies have started to revolutionize both clinical and cognitive neuroscience.

The availability of such new devices made possible by pioneering breakthroughs in physics and engineering now pose major computational and statistical challenges for which machine learning currently plays a major role. In this course you will discover hands-on the types of data one can collect to record the living brain. Then you will learn about state-of-the-art supervised machine learning approaches for EEG signals in the clinical context of sleep stage classification as well as brain computer interfaces. ML techniques that will be explored are based on deep learning as well as Riemannian geometry that has proven very powerful to classify EEG data. You will do so with MNE-Python (<https://mne.tools>) which has become a reference tool to process MEG/EEG/sEEG/ECOG data in Python, as well as the scikit-learn library (<https://scikit-learn.org>) based on PyTorch. The teaching will be done hands-on using Jupyter notebooks and public datasets, that you will be able to work using google collab.

Finally, this tutorial will be a unique opportunity to see what ML can offer beyond standard applications like computer vision, speech or NLP.

Tutorial 3B – Hybrid Artificial Intelligence and Image Understanding

Speaker: Isabelle BLOCH

Abstract: The tutorial will review a few methods for symbolic AI, for knowledge representation and reasoning, and show how they can be combined with learning approaches for image understanding. Examples in medical image understanding will illustrate the talk.

Tutorial 4B – An Introduction to Deep Reinforcement Learning

Speaker: Corentin TALLEC

Abstract: Be it on Atari Games, Go, Chess, Starcraft II or Dota, Deep Reinforcement Learning (DRL) has opened up Reinforcement Learning to a variety of large-scale applications. While it could formally appear as a straightforward extension of reinforcement learning to deep learning based function approximations, DRL often involves more than simply plugging the newest deep learning architecture into the best theoretical reinforcement learning method. In this tutorial, we will journey through the recent history of DRL, from the now seminal Neural fitted-Q to the most popular Deep Q-Network (DQN). Alongside the lecture, the practical session will revolve around implementing and testing DRL algorithms in JAX and Haiku on simple environments.

Tutorial 5B – Deep Learning for Audio Processing

Speaker: Geoffroy PEETERS

Abstract: As in many fields, deep neural networks have allowed important advances in the processing of audio signals. In this tutorial, we review the specificities of these signals, elements of audio signal processing (as used in the traditional machine-learning approach) and how deep neural networks (in particular convolutional ones) can be used to perform feature learning (without prior knowledge -- 1Dconv, TCN--, or using prior knowledge --source/filter, auto-regressive, HCQT, SincNet, DDSP--). We then review the dominant DL architectures, meta-architectures and training paradigms (classification, metric learning, supervised, unsupervised, self-supervised, semi-supervised) used in audio. We exemplify the used of those for some key applications in music and environmental sounds processing: sound event detection, localization, auto-tagging, source separation, generation.

Tutorial 6B – Reliable Decision Making and Causal Inference with Kernels

Speaker: Krikamol MUANDET

Abstract: Data-driven decision-making tools have become increasingly prevalent in society today with applications in critical areas like health care, economics, education, and the justice system. To ensure reliable decisions, it is essential that the models learn from data the genuine correlations (i.e., causal relationships) between the outcomes and the decision variables. In this tutorial, I will first give an introduction to the causal inference problem from a machine learning perspective including causal discovery, treatment effect estimation, instrumental variable (IV), and proxy variables. Then, I will review recent developments in how we can leverage machine learning (ML) based methods, especially modern kernel methods, to tackle some of these problems.

Round Tables

Industry Round Table

The Industry Panel, composed of Hi! PARIS Corporate Donors’ representatives is an opportunity to talk about and disseminate their activities in AI/Data Science and also start an interchange with the academic community. The rond table will be moderated by our scientific committee member Prof. Jean-Edouard Colliard (HEC Paris).

After an opening introduction by the moderator, each of the panel members will be asked to present their latest AI initiatives, opportunities, and challenges.

The discussion will be led by the moderator with the speakers on and questions from the audience.

We hope the industry panel will be a very interactive event with an opportunity to open communication channels for further research opportunities between the industry and the academic community.

Academic Round Table

The panel is formed of our dear keynote speakers and will be moderated by our scientific committee member Prof. Anna Korba (IP PARIS – ENSAE Paris).

After an opening introduction, each of the panel members will be asked to give a very short presentation of their research (about 5-minutes).

The discussion will be led by the moderator with the speakers on personal point of view of the speakers about priorities in AI research, evolution of the field of AI, personal evolution of thinking/using AI, academic and industrial job markets. Will follow questions from the audience.

Student Program

Who Better To Give Professional Advice Than Students Who Have Gone Through The Same Situations As You?

The student program is a conference round table organized by our students, Willie Hernandez and Aurore Troussel (HEC Paris).

This activity aims to introduce Brayam, Ahmed and Dilia, former students of Institut Polytechnique de Paris who are currently working in the Data Science in the industry. In this opportunity, we will ask them about their experiences at the university and in their work. They will present their tips on how to get a job, they will tell us the Do's and Don't's when presenting a job interview and finally, they will tell us how their experience at IP helped them to become the successful professionals they are today.

Who better to give professional advice than students who have gone through the same situations as you?

Speakers:

Ahmed LACHTAR – Ernst & Young, graduate of ENSTA Paris
Brayam VELANDIA – Technopolis Group, graduate of Telecom Paris
Dilia OLIVO – La Banque Postale, PhD at Telecom Paris

Social Time

Conference Networking And Coffee

is an opportunity for participants in the summer school to interact with others and chat over a drink or a cup of coffee.

Social Event

Great time evening event at [AllTogether Saclay](#), a leisure student bar close to Télécom Paris. Attendees to the Summer School will be able to enjoy around games, drinks, and finger food!