

# Activity Highlights

5<sup>th</sup> MARCH 2020



# Table of Contents

---

<b>Message from the Director</b>	<b>5</b>
<b>The DATAIA Institute</b>	<b>6</b>
Introduction	7
Executive Board	8
DATAIA Support Team	9
Scientific Advisory Board	10
Steering Board	11
Saclay AI Chairs	12
<b>Research Activity</b>	<b>14</b>
Research Projects	15
PhD Thesis	19
Invited Professors	19
<b>Community Animation</b>	<b>20</b>
Recurrent Events	21
Major Joint Events	22



# Message from the Director

---

“The mission of the Convergence Institute DATAIA is to federate the academic initiatives of the Saclay campus in artificial intelligence (AI), data science and societal impacts.

The diversity and the influence of Saclay's ecosystem are unique in France. These assets make it the leading AI research community in Europe. In April 2019, Florence Parly, the French Minister of Army, visited the DATAIA Institute. She reminded us, on one hand, of the importance of the site for the national AI strategy, in a context of fierce competition between academic and industrial players; and on the other hand, of the challenges in safety and AI.

I have agreed to take over from Nozha Boujemaa because I believe in the strength of the scientific community in Saclay, that is why I will strive to develop more interactions within our ecosystem.



*Florence Parly, French Minister of Army, visiting the DATAIA Institute in April 2019*



To this end, we have launched events in series (seminars and workshops), and we continue DATAIA's successful actions such as project calls and international symposia.

This booklet will give an overview of the activities we have carried out since our kick-off in 2018.

One of highlights that we want to share with you is the result of the national call «Research and teaching chairs in artificial intelligence»: 16 out of the 40 projects selected are led by researchers of the Saclay campus. They will have a structuring effect on research, and consequently on trainings and innovation.

The DATAIA Institute will continue to evolve, along with the academic players of the Saclay campus and its environment. Our three ongoing projects in 2020 are: having strategic foresights, finding new ways of interacting with the industrial world, and developing a simplified governance. Your proposals and suggestions are welcome: DATAIA is what we all make of it.

**Bertrand Thirion**  
Director of the DATAIA Institute

# THE DATAIA INSTITUTE

---

**Introduction**

**Executive Board**

**DATAIA Support Team**

**Scientific Advisory Board**

**Steering Board**

**Saclay AI Chairs**

# Introduction

Created in 2017 as part of the national «Call for Convergence Institutes» launched by the *Agence Nationale de la Recherche (ANR)*, the DATAIA Institute is the first French ecosystem in artificial intelligence and data science.

Carried by Paris-Saclay University, and under the scientific supervision of Inria, it gathers the multidisciplinary expertise of 17 higher education and research organisations in Saclay.

Its main ambition is to support its partners to become major actors in cutting-edge research in AI, data science, and human and social sciences.

Since its launch, it has contributed to innovation by carrying out many actions: seven annual and permanent calls, organisation and support of more than 40 events such as seminars, conferences, data challenges and workshops.



17

academic partners



11

funded research projects



56

partner laboratories



1,5K

researchers

# Executive Board

The Executive Board (*Bureau*) is composed of ten experts from the DATAIA Institute's partners, who are in charge of a specific mission.

The board manages the every day

business of the Institute and runs the scientific and strategic actions under the Supervision Board (*Comité des Tutelles*) and the Steering Board (*Comité des Programmes*).



## **Bertrand Thirion - Director of the DATAIA Institute**

Bertrand is head of the Parietal team at the Inria-Saclay center. He is a contributor of statistics and machine learning methods, and software for brain imaging (scikit-learn, nilearn).

## **Christine Balagué in charge of Observatory of Use Cases**

Christine is Professor and holder of the Good in Tech Chair at Institut Mines-Télécom Business School. She is also member of the Board of Directors of Cap Digital and of the CSA expert committee.



## **Alexandra Bensamoun in charge of Regulation of AI**

Alexandra is Professor of law at the University of Rennes 1/University of Paris-Saclay, and IP/IT expert. She is a «qualified personality» for the CSPLA and submitted a report on AI and Culture in 2020.

## **Sarah Cohen Boulakia in charge of Trainings at Paris-Saclay University**

Sarah is Professor at the *Laboratoire de Recherche en Informatique* (LRI) lab. She is strongly involved in the Center for Data Science and director of the GDR MaDICS of CNRS on data sciences.



## **Bruno Defude in charge of DATAIA Institute Ecosystem**

Bruno is Professor at Télécom Paris and researcher in the SAMOVAR laboratory. He is deputy scientific manager of the laboratory of excellence in ICST (labEx DigiCosme) on the Saclay campus.

**Laurence Devillers** in charge of **International Relations**

Laurence is Professor at the Paris-Sorbonne University and researcher at the LIMSI of the CNRS. In 2020 she has been nominated *Chevalière de la Légion d'honneur*.



**Frédéric Pascal** in charge of **Operational Coordination of Programs**

Frédéric is a full Professor in the L2S laboratory at CentraleSupélec. He is also the coordinator of AI activities at CentraleSupélec and holder of the Givaudan Chair on data sciences.

**Jean-Noël Patillon** in charge of **Industrial Relations**

Jean-Noël contributes to the setting up of the Nano-INNOV operation and the SystemX Technological Research Institute. He is acting director of the CEA List and assistant to the director of the CEA List Carnot Institute.



**Michèle Sebag** in charge of **Research**

Michèle, co-founder and head of the machine learning team TAU (CNRS - Inria - LRI), is principal scientist at the CNRS, and member of the *Académie des Technologies* and *Chevalière de la Légion d'Honneur*.

**Fabian Suchanek** in charge of **Trainings at Institut Polytechnique de Paris**

Fabian is Professor at Télécom Paris. He has developed the YAGO knowledge base, one of the most important public knowledge bases for general use.



## DATAIA Support Team



**Viviane Hoang**  
Project Manager



**Jasmyn Scaramella**  
Communications  
Officer



**Eric Tordjeman**  
Industrial Partnerships  
Manager

# Scientific Advisory Board

The Scientific Advisory Board assesses and directs the actions of the DATAIA Institute.

It provides guidance on the main orientations of the Institute's scientific policy and it ensures coherence

between the evolution of the activity and the scientific programs.

The committee is composed of top-level international renown scientists, who are not affiliated with any of the Institute's partners.



## **Masaru Kitsuregawa- Professor at University of Tokyo (Japan)**

Masaru, Kitsuregawa is Director of the National Institute of Informatics. He has wide research interests, especially in database engineering. He has received many awards including ACM SIGMOD E. F. Codd Innovations Award, IEICE Contribution Award, IPSJ Contribution Award, 21st Century Invention Award of National Commendation for Invention, Japan and C&C Prize. In 2013, he was rewarded with the Purple Ribbon medal and in 2016, *Chevalier de la Legion D'Honneur*. He is a fellow of ACM, IEEE, IEICE and IPSJ.

## **François Laviolette - Professor at University of Laval (Canada)**

Professor François Laviolette is the founding Director of Laval University's Big Data Research Center (BDRC).

He is a leader in the Bayesian PAC theory, and his research focuses on artificial intelligence, especially machine learning. Part of his research is to extend interpretability of AIs to other areas, such as insurance. In this regard, he owns two research chairs: one from NSERC and the second one from CIFAR.



## **Verena Rieser - Professor at Heriot-Watt University (UK)**

Verena Rieser is affiliated with the Interaction Lab at Heriot-Watt University, and the Edinburgh Centre for Robotics. Her research focuses on machine learning techniques for conversational AI. She has served the research community in a number of leadership roles, as well as scientific advisor to the government and public bodies.

For the past two years, Verena and her group were the only UK team to make it through to the finals of the Amazon Alexa Prize.

### **John Shawe-Taylor - Professor at University College London (UK)**

John Shawe-Taylor has helped to drive a fundamental rebirth in the field of machine learning, with applications in novel domains such as computer vision and document classification. He has also been instrumental in assembling a series of influential European Networks of Excellence.

He was appointed UNESCO Chair of Artificial Intelligence in November 2018 and he is the leading trustee of the UK Charity, Knowledge 4 All Foundation.



### **Henri Verdier - Ambassador for digital affairs (France)**

A former student of the École Normale Supérieure, Henri Verdier was the founder and CEO of the company Odile Jacob Multimedia.

He is one of the founding members of the Cap Digital competitiveness cluster, and he served as Vice-Chairman from 2006 to 2008, before becoming Chairman of the Board of Directors from 2008 to January 2013. Henri Verdier is also a member of the Scientific Council of the Institut Mines-Télécom, as well as a member of the ARCEP Foresight Committee and the CNIL Foresight Committee.



### **Stephan Wrobel - Professor at University of Bonn (Germany)**

Professor Stefan Wrobel is Director of the Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS and one of the directors of the Bonn-Aachen International Center for Information Technology (b-it).

His work is focused on intelligent systems and their productive use in business applications.

He is on the Editorial Board of several leading academic journals in his field, co-speaker of ML2R, and speaker of the Fraunhofer Alliance on Big Data and AI.



## **Steering Board**

The Steering Board (*Comité des Programmes*) is currently composed of scientific representatives from the founding institutions and representatives from the DATAIA Institute's ecosystem: coordinators of the objects of excellence in Saclay, with which the Institute has common activities, and heads of the Math, STIC, and SHS

departments of the Paris-Saclay University.

The Steering board evaluates and ranks scientific projects and actions submitted to the Institute.

The scientific representatives of the board vote the financing of initiatives related to research and training activities within the DATAIA Institute.

# Saclay AI Chairs

The call for «Research and teaching chairs in artificial intelligence» aims to offer French and foreign researchers, with the support of their host institutions, resources to build a team and carry out

an ambitious, high-impact project.

Among the 40 selected projects countrywide, 16 are led by researchers who are involved in the activity of the DATAIA Institute.



**Gilles Blanchard (Paris-Saclay University)**  
BISCOTTE *#Statistics #Efficiency*



**David Bounie (Télécom Paris)**  
XAIforAML *#Explainability #Safety*



**Frédéric Chazal (Inria)**  
TopAI *#TopologicalDataAnalysis*



**Laurence Devillers (Paris-Sorbonne University, CNRS)**  
HUMAINE *#HumanMachine #Emotions #Ethics*



**Edouard Duchesnay (CEA)**  
Big2small *#TransferLearning #Neuroimaging*



**Alexandre Gramfort (Inria)**  
BrAIN *#Neuroscience #Optimization*



**Isabelle Guyon (Paris-Saclay University)**  
HUMANIA *#Democratization #AIforGood*

**Jean-François Mangin (CEA)**  
FOLDDICO *#Neuroscience*



**Ioana Manolescu (Inria, École polytechnique)**  
SourcesSay *#DataIntegration #DataExploration*

**Éric Moulines (École polytechnique)**  
SCAI *#DeepLearning #BayesianInference*



**Maks Ovsjanikov (École polytechnique)**  
AIGRETTE *#3DDeepLearning #ShapeAnalysis*

**Jean-Christophe Pesquet (CentraleSupélec)**  
BRIDGEABLE *#Optimization #SignalProcessing*



**Fabian Suchanek (Télécom Paris)**  
NoRDF *#InformationModeling #InformationExtraction*

**Bertrand Thirion (Inria)**  
KARAIB *#HeterogeneousData #Knowledge*



**Gaël Varoquaux (Inria)**  
LearnI *#DataIntegration #DatabaseLearning*

**Michalis Vazirgiannis (École polytechnique)**  
AML-HELAS *#LargeScaleData #HeterogeneousData*



# RESEARCH ACTIVITY

---

Research Projects

PhD Thesis

Invited Professors

# Research Projects

The DATAIA Institute supports multidisciplinary research through an annual call for research projects.

11 research projects have been funded so far. In 2020, three research projects will be funded: one fully by the

DATAIA Institute, one with MSH Paris-Saclay and the third one, in collaboration with the DigiCosme Labex. The goal is to encourage collaborations between researchers from different scientific fields.



## Bad Nudge - Bad Robot?

Connected devices, specifically conversational agents such as Google Home or social robots, bring speech as a new dimension to interaction. Users tend to anthropomorphize these devices which could be soon able to detect user's emotions and then will become a means of influencing individuals. They are currently neither regulated, nor evaluated and very opaque.

This project is studying nudges in human-machine verbal interaction, to understand the societal impact of these new devices. The ultimate objective of the project is to create «ethic-by-design» systems and also to create metrics of evaluation. This project is linked to the AI chair HUMAINE.

**Funded:** 2018

**Project directors:** Laurence Devillers (LIMSI-CNRS, Paris-Sorbonne University), Serge Pajak (RITM, Paris-Saclay University)

## GDP - ERE

Within the framework of Personal Cloud tools, depending on the architectures, the user who is qualified as the controller of his data.

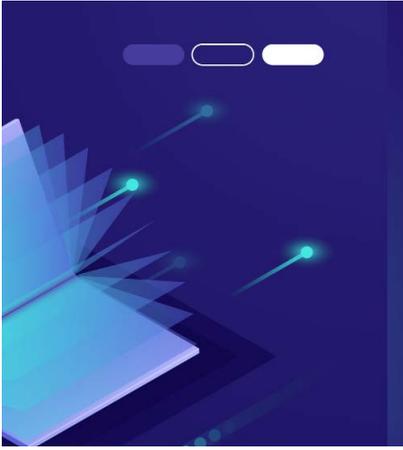
This research project focuses on the distribution of responsibility between the user and the supplier in Personal Cloud architectures.

Its objectives are the following: to analyze the impact of current Personal Cloud architectures on user responsibility and compare this analysis with the legislation and rules laid down by the RGPD; and to formulate legislative and technological recommendations in order to preserve the autonomy of the user.



**Funded:** 2018

**Project directors:** Nicolas Anciaux (Inria), Mélanie Clément-Fontaine (UVSQ), Philippe Pucheral, Guillaume Scerri (UVSQ, Inria), Célia Zolynski (University Paris 1 Panthéon-Sorbonne)



## HistorIA

This project brings together researchers in history, computational social sciences and information visualization.

The goal is to develop and explore large historical databases by applying data mining methods supported by visualization, while implementing an iterative approach to the exploration process, based on users' appropriation of the procedures, tools used, and the results of the analyses.

To this end, emphasis will be placed on the explainability of the algorithms, on progressive data analysis, and human-machine interaction.

**Funded:** 2018

**Project directors:** Jean-Daniel Fekete (Inria), Christophe Prieur (Télécom Paris)

## Information Leakage in Deep Learning

This research project aims to propose an analysis of machine learning models to uncover and quantify information leakage in deep learning. By leveraging recent results, tools will be developed and used to analyze white and black-box threat models.

Focus will be put on the analysis of the state-of-the-art attacks to privacy in learning systems, the amount of sensitive information which can be retrieved from the trained software, and the possible strategies to reduce the privacy threats.



**Funded:** 2019

**Project directors:** Catuscia Palamidessi (Inria, École polytechnique), Georg Pichler (TU Wien), Pablo Piantanida (CentraleSupélec), Marco Romanelli (Inria, École polytechnique)



## MissingBigData

Many current databases contain missing data, resulting for instance from the merging of different sources, missing information, or input errors.

The MissingDataIA project aims to develop a framework for the analysis of predictive algorithms for machine learning in the presence of missing data.

The project is working on health data but the ambition is to produce a generic model, and a method applicable to other fields as well.

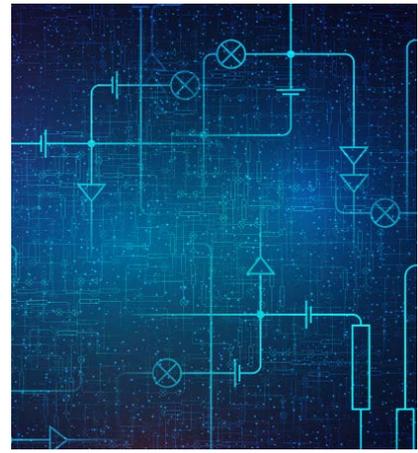
**Funded:** 2018

**Project directors:** Julie Josse (École polytechnique, Inria), Gaël Varoquaux (Inria)

## PEPER

How to make production, consumption and storage cooperate for a better use of renewable energies?

The PEPER project is studying and developing three concepts in order to create a balanced system for the efficient management of renewable energies: production, consumption, and storage. The objective is to collect data on the different actors of this network, and to use learning and deep reinforcement learning techniques to develop algorithms to predict the production and consumption of each actor, and then to allow cooperation between them.



**Funded:** 2018

**Project directors:** Hossam Afifi (Télécom Paris), Jordi Badosa (École polytechnique), Florence Ossart (CentraleSupélec)



## Smart Lawyer

The expansion and globalization of legal services contrast sharply with the lack of sophistication of its measurement tools and the strong information asymmetry between providers and consumers of those services.

The objective of the project is to automate the retrieval of legal and factual information from court decisions in selected jurisdictions and produce analytics of legal services in the courtroom. More specifically, the project will generate, among others, an original dataset of annotated court decisions, social networks of actors, and metrics of services. For this purpose it focuses in the fields of civil and commercial law.

**Funded:** 2018

**Project directors:** David Restrepo-Amariles (HEC Paris), Michalis Vazirgiannis (École polytechnique)

## StreamOps

This research project aims at creating a novel approach, and tool to develop powerful algorithms capable of managing and mining data flows.

It aspires to be at the interface of algorithmic, business and software aspects to offer both researchers and engineers a generic data stream processing and machine learning platform, at the forefront of algorithms through: easy integration of new algorithms, operational robustness, detection and information compression performance, and consideration of data confidentiality and problems related to real data.



**Funded:** 2018

**Project directors:** Marc Fischler (Hôpital Foch, UVSQ), Cédric Gouy-Pailler (CEA), Yehia Taher (UVSQ), Karine Zeitouni (UVSQ)



## UltraBioLearn

Machine learning for medical applications, has many constraints related to results interpretability, quality, trustworthiness, and more. At the same time, the private lives of patients must be respected. That is why access to medical data is restricted.

This project proposes to research innovative solutions to these issues, in particular by leveraging semi-supervised learning using generative, graph-based and certifiable networks, in the context of predicting patient response to cancer treatments.

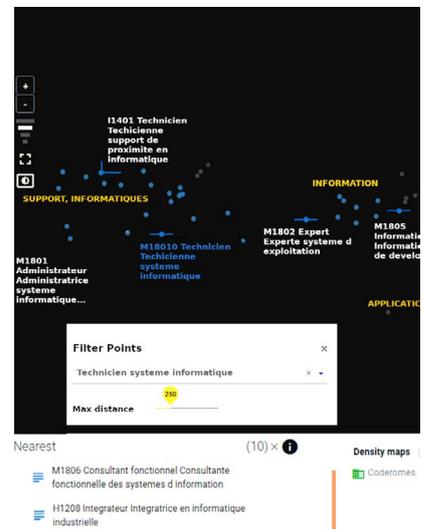
**Funded:** 2019

**Project directors:** Caroline Caramella, Nathalie Lassau (CNRS, Paris-Saclay University), Younes Belkouchi, Fragkikos Malliaros, Hugues Talbot (CentraleSupélec)

## Vadore

In partnership with Pôle Emploi, Vadore investigates the recommendation of employment opportunities to job seekers and vice versa: a matching function, taking into account potential market congestion, is learned by combining collaborative filtering and optimal transport. It will be assessed by randomized control trial in 2020.

The challenges addressed in Vadore include: identifying the biases in the data; defining performance indicators at the population level; proposing an interactive representation of the employment opportunities; relying on a topology of the work space; and supporting the job seekers and recruiters' empowerment.



**Funded:** 2018

**Project directors:** Philippe Caillou (LRI), Bruno Crepon, Christophe Gaillac (ENSAE), Marco Cuturi (Google Brain, ENSAE), Michèle Sebag (CNRS, Paris-Saclay University)



## Warm Rules

The project aims at developing an explainable approach to discover gradual causal rules in knowledge graphs. The main challenges are: (i) to develop a new explainable and efficient approach for the discovery of causal gradual rules based on ontology semantics, statistics, temporal information and contextual identity; (ii) to define new measures for quality assessment and ranking of causal gradual rules; and (iii) to build a new knowledge graph representing relations between global warming and corn phenotype development.

**Funded:** 2019

**Project directors:** Juliette Dibie (INRA, AgroParisTech), Elodie Marchadier (GQE, Paris-Saclay University), Fatiha Saïd (LRI, Paris-Saclay University)

# PhD Thesis

## Artificial intelligence, automation and political action

**François Levin**  
(LinX, École polytechnique)

### Thesis supervisors:

Jean-Gabriel Ganascia (Paris 6),  
Michaël Foessel (École polytechnique)

The development of connectionist artificial intelligence in the past years, and specifically machine learning, has brought a flow of criticism about the political impacts of these technologies.

Increase of automation is seen as the principal problem raised by artificial intelligence.

The objective of my work will be first to question this notion of automation: what does it mean? What do we gain from automating? What is endangered by automation?

I will then try to determine how to act politically in a context of automation: should we rebel against automation or can we turn it against itself in order to create new opportunities?

## Deep learning for sets with graphs

**Moussa Kamal-Eddine**  
(Paris-Saclay University)

### Thesis supervisors:

Fragkiskos Malliaros (CentraleSupélec, Inria),  
Michalis Vazirgiannis (École polytechnique)

In this PhD we aim to tackle the problem of abstractive summary generation. Our main goal is to define a neural based model that is able to generate good quality summaries.

Abstractive summarization poses many challenges, some of them are related to the quality of the generated summaries and others are related to their evaluation. We want to tackle them by proposing novel and efficient neural network architectures, and using different transfer learning and pretraining techniques. Furthermore, we will work on finding new automatic evaluation metrics that go beyond simple n-grams overlapping techniques such as ROUGE and BLEU metrics.

# Invited Professors



**Aapo Hyvärinen**  
University of Helsinki



**Matthieu Jonckheere**  
University of Buenos Aires



**Valderio Reisen**  
Federal University of  
Espírito Santo

# COMMUNITY ANIMATION

---

**Recurrent Events**

**Major Joint Events**

# Recurrent Events

## DATAIA Seminars

Local and international guests are invited every month to discuss with the Saclay community about their research work.



## DATAIA Workshops

Experts and professionals from the concerned disciplines are gathered to share knowledge and challenges, consolidate a locally rooted research community and pave the way for the construction of joint projects.



## DS3

Co-organised by École polytechnique and the DATAIA Institute, the Data Science Summer School aims to lively share and exchange on the latest advances in the field, with internationally renowned professors and experts of the area.

## International Symposia

The ongoing dialogue between France and foreign research communities is enriched with international symposia that are organised every year.

The DATAIA Institute holds proven relationships with IVADO (Canada), the Alan Turing Institute (UK), and NII & JST (Japan).



# Major Joint Events



JSDE 2018 - 2019

SCI-KIT LEARN SPRINT  
2019



MLAI 2019

LE SÉMINAIRE PALAISIEEN



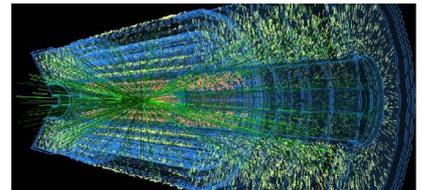
MEET-UP 2019  
DOCTORANTS & INDUSTRIE

CONFÉRENCE  
« IA ET CULTURE »



BACTERIA BIOTOPE  
CHALLENGE

TRACK ML CHALLENGE



The DATAIA Institute calls «Call for workshop and conferences» and «Support for the organization of Data

Challenges» are permanently open. For more information: [dataia.eu/en/our-calls](http://dataia.eu/en/our-calls)



INSTITUTE  
**DATAIA**  
Data Science, Intelligence & Society



Institut Convergence 17-CONV-0003 INSTITUT DATAIA (I2DRIVE)

**Institut DATAIA**

Centre de Recherche Inria Saclay - Île-de-France  
Campus de l'Ecole Polytechnique - Bâtiment Alan Turing  
1 rue Honoré d'Estienne d'Orves  
91120 Palaiseau

**Service communication**  
com-dataia@inria.fr



[www.dataia.eu](http://www.dataia.eu)



[@institut\\_dataia](https://twitter.com/institut_dataia)