

# Call for Master 2 internships 2024

## Introduction

The DATAIA Institute is the French convergence institute dedicated to data science, artificial intelligence and their impact on society. It is at the crossroads of Big Data technologies and artificial intelligence with other disciplines, notably the human and social sciences, for artificial intelligence at the service of human beings. Selected by the French National Research Agency (ANR) as part of the "Programmes d'Investissements d'Avenir", the DATAIA Institute aims to bring together and structure multi-disciplinary expertise to better respond to the major challenges of data science, artificial intelligence and their applications.

The institute supports multi-disciplinary research in Data Science in the broadest sense - including computer science, mathematics, physical sciences, life sciences and human and social sciences - through an annual call for Master 2 internships.

## Prerequisites

The project leader must belong to a laboratory of a DATAIA partner institution listed in Appendix 2 of the call. The proposal must fall within the scope of the DATAIA issues described in appendix 1. Projects involving several laboratories from different DATAIA partner establishments will be highly appreciated.

## Important dates

<b>Closing</b>	November 10th, 2023 (12pm)
<b>Decision</b>	December 8th, 2023
<b>Contact</b>	contact-dataia@inria.fr

## Submission procedure

The application must include:

### Mandatory information:

- Project title;
- Names of supervisors and their contact details;
- Names of host laboratories and/or research teams

**Project description (2 pages):**

The partners describe the objective of the internship, its link with DATAIA, and list the partners' previous collaborations (if any). It is not essential to have a candidate to submit the project.

**Financial appendix indicating:**

- Amount of assistance requested and duration of internship;
- Planned start and end dates of internship;
- Name of the institution receiving and managing the funds;
- Names and contact details of administrative and financial contacts in charge of project follow-up.

**A template is available at the following address:**

<https://www.dataia.eu/sites/default/files/2024-05/Template%20AAP%20M2-vEN.pdf>

The application must be sent by the host laboratory to: [submission-dataia@inria.fr](mailto:submission-dataia@inria.fr)

**Support:**

If the project is accepted, DATAIA will fund the student for 4 to 6 months at the usual stipend level (around 650 euros per month). For further information:

- The student must be present for more than 308 hours during the same teaching year to qualify for a bonus;
- Public bodies may not pay more than the minimum legal amount for an internship, on pain of the internship agreement being requalified as an employment contract;
- The legal minimum amount in 2023 for an internship bonus is €4.05 net per hour, with no charges to be paid;
- The hourly ceiling for 2024 will not be known until the decree is published in the Journal Officiel in November or December 2023.

**All communications (publications, presentations) must mention  
DATAIA funding:**

*“This research was supported by DATAIA convergence institute as part of the “Programme d'Investissement d'Avenir”, (ANR-17-CONV-0003) operated by [Partner] XXX”*

## Appendix 1: DATAIA research areas

The scientific program of each proposal must explicitly address at least one of the four research areas outlined below, which are DATAIA's four interdisciplinary challenges.

### **FROM DATA TO KNOWLEDGE, FROM DATA TO DECISIONS**

The growing availability of massive data is pushing back the technical frontiers in many fields. On the one hand, the heterogeneous, semi-structured, incomplete or uncertain nature of data calls into question the usual statistical models and algorithms dedicated to decision-making. On the other hand, data management raises new operability constraints, such as security, integrity and traceability.

What's more, producing knowledge requires building models that deliver explainable, statistically valid and calculable decisions. Acceptance of results also requires that confidentiality and loyalty be reinforced. At the same time, new developments in optimization should make it possible to improve estimation procedures.

Challenges:

- Heterogeneous, complex, incomplete, semi-structured and/or unsure data
- Massive data: algorithms and data structuring
- E-learning, methodology for massive data, efficient methods
- Improved storage, computation and estimation for data science
- Game-theoretic modeling of interactions between agents (human or artificial)
- Multiscale and multimodal representation and algorithms
- Theoretical analysis of heuristic methods (complexity theory, information geometry, Markov chain theory)
- Human-machine coevolution in autonomous systems: conversational agents, cars, social robots

### **LEARNING AND ARTIFICIAL INTELLIGENCE**

Recently, deep learning research has made spectacular advances in computer vision and natural language processing. Beyond the arrival of massive data, increased computing power and design efforts, the causes of these advances, which are still poorly understood, raise at least three questions. What learning theory will enable us to analyze deep architectures? How can we manage the compositionality of these architectures and their ability to apprehend more complex objects? How can we open the black box to update learned representations?

Challenges:

- Innovative machine learning and AI: common sense, adaptability, generalization
- Deep learning and adversarial learning
- Machine learning and hyper-optimization
- Optimization for learning (e.g. improvements in stochastic gradient methods, Bayesian optimization), combinatorial optimization
- Learning-modeling link, a priori integration in learning
- Reproducibility and robust learning
- Statistical inference and validation
- Compositionality of deep architectures.

### **TRANSPARENCY, RESPONSIBLE IA AND ETHICS**

Digital trust is built on the implementation of ethically responsible methodologies through the transparency and accountability of algorithmic systems; the regulation of the collection, use and processing of personal data; and the reinforcement of regulation through appropriate digital procedures. Privacy by design is a form of regulation that includes the protection of personal data at all stages of collection and processing. The tracing of tools applied to data must also be developed in such a way as to facilitate the explanation of the model for experts and users alike, making algorithmic systems auditable. Confidentiality principles, although easy to formulate, require modifications to storage and processing infrastructures, with major legislative, sociological and economic impacts. Transparency techniques for algorithmic systems will be developed, focusing on: fairness, loyalty and non-discrimination, and accountability-by-construction.

Challenges:

- Accountability and explicability by design
- Transparency and fairness by design
- Auditing algorithmic systems: non-discrimination, fairness, technical bias, neutrality, equity
- Measuring trust and digital appropriation
- Progressive user-analysis of progressive data (interactive monitoring of decision-making systems)
- Responsibility for information processing and decision-making: data usage control and fact-checking
- Causal discovery, traceability of inferences from source data, interpretability of deep architectures

### **PROTECTION, REGULATION AND THE DATA ECONOMY**

Companies involved in the data economy continually need to rethink how they are structured: they must adopt a project-oriented organization with rapid changes in resource allocation. The data economy also raises issues of concentration and monopoly. A small number of companies (GAFAM) hold most of the data. This market concentration can lead to unfair competition, and innovation in small and medium-sized businesses is likely to suffer. Citizens expect governments to intervene in the digital economy to prevent too much concentration and monopoly. Governments must prevent information leakage to preserve state sovereignty and respect for regulations.

Challenges:

- Privacy by GDPR design
- Privacy-aware learning (differential privacy)
- Development of ethically responsible methodologies and technologies to regulate the collection, use and processing of personal data, and the exploitation of knowledge derived from such data
- IT security for data processing chains
- Security/crypto: blockchain and trusted third parties

Projects may address topics more specific to particular fields of application. A list of these fields is given below as an indication. However, it is requested that projects address data science issues of general interest, i.e. not specifically restricted to one application area.

- *Energy: optimization of management (production and distribution), and regulation*
- *Health, personalized nutrition and well-being*
- *Urbanization Mobility (connected and autonomous vehicles, smart cities, etc.)*
- *Analysis for finance and insurance*
- *Internet of people and things*
- *E-Sciences*

## Appendix 2: List of DATAIA partners eligible for the call for proposals

<ul style="list-style-type: none"><li>• Agro ParisTech</li><li>• CentraleSupélec</li><li>• CEA</li><li>• CNRS</li><li>• ENS Paris-Saclay</li><li>• IFP- Énergies Nouvelles</li><li>• Inria</li></ul>	<ul style="list-style-type: none"><li>• Institut Mines-Télécom Business School</li><li>• Inrae</li><li>• ONERA</li><li>• Université d'Evry Val d'Essonne</li><li>• Université Paris-Saclay</li><li>• Université Versailles St-Quentin-en-Yvelines</li></ul>
--	---