

RESEARCH CENTRE

**Inria Centre at Rennes  
University**

IN PARTNERSHIP WITH:

Institut national des sciences appliquées  
de Rennes, Institut national supérieur des  
sciences agronomiques, agroalimentaires,  
horticoles et du paysage, Université de  
Rennes

2024

ACTIVITY REPORT

Project-Team  
**LACODAM**

## **Large scale Collaborative Data Mining**

IN COLLABORATION WITH: Institut de recherche en informatique et  
systèmes aléatoires (IRISA)

**DOMAIN**

**Perception, Cognition and Interaction**

**THEME**

**Data and Knowledge Representation and  
Processing**

*Inria*

# Contents

<b>Project-Team LACODAM</b>	<b>1</b>
<b>1 Team members, visitors, external collaborators</b>	<b>3</b>
<b>2 Overall objectives</b>	<b>4</b>
<b>3 Research program</b>	<b>4</b>
3.1 Introduction	4
3.2 Symbolic methods	5
3.3 Interpretable ML	5
3.4 Real world AI	6
<b>4 Application domains</b>	<b>6</b>
4.1 Industry	6
4.2 Agriculture and Environment	7
4.3 Cognitive Sciences	8
4.4 Semantic Data Management	9
<b>5 Social and environmental responsibility</b>	<b>9</b>
5.1 Footprint of research activities	9
5.2 Impact of research results	9
<b>6 Highlights of the year</b>	<b>10</b>
6.1 Awards	10
<b>7 New software, platforms, open data</b>	<b>10</b>
7.1 New software	10
7.1.1 HIPAR	10
7.1.2 Dexteris	10
7.1.3 skm	11
<b>8 New results</b>	<b>11</b>
8.1 Symbolic Methods	11
8.1.1 Pattern Mining	11
8.1.2 Graph-FCA	12
8.1.3 Semantic Web	12
8.2 Interpretable Machine Learning	13
8.3 Real World AI	15
8.3.1 Computer Vision and Robotics	15
8.3.2 Agriculture	16
8.3.3 Machine Learning on Sequences	17
<b>9 Bilateral contracts and grants with industry</b>	<b>18</b>
9.1 Bilateral contracts with industry	18
<b>10 Partnerships and cooperations</b>	<b>19</b>
10.1 International research visitors	19
10.1.1 Visits of international scientists	19
10.2 European initiatives	20
10.2.1 Horizon Europe	20
10.2.2 H2020 projects	20
10.3 National initiatives	20
10.3.1 ANR	22

<b>11 Dissemination</b>	<b>24</b>
11.1 Promoting scientific activities	24
11.1.1 Scientific events: organisation	24
11.1.2 Scientific events: selection	24
11.1.3 Journal	25
11.1.4 Invited talks	25
11.1.5 Leadership within the scientific community	25
11.1.6 Scientific expertise	26
11.1.7 Research administration	26
11.2 Teaching - Supervision - Juries	26
11.2.1 Teaching	26
11.2.2 Supervision	28
11.2.3 PHD & HDR Juries	29
11.2.4 Doctoral advisory comitee (CSID)	30
11.3 Popularization	30
11.3.1 Specific official responsibilities in science outreach structures	30
11.3.2 Productions (articles, videos, podcasts, serious games, ...)	30
11.3.3 Participation in Live events	30
11.3.4 Others science outreach relevant activities	31
<b>12 Scientific production</b>	<b>31</b>
12.1 Major publications	31
12.2 Publications of the year	32
12.3 Cited publications	35

## Project-Team LACODAM

*Creation of the Project-Team: 2017 November 01*

### Keywords

#### Computer sciences and digital sciences

- A2.1.5. – Constraint programming
- A3.1.1. – Modeling, representation
- A3.1.2. – Data management, quering and storage
- A3.1.6. – Query optimization
- A3.1.11. – Structured data
- A3.2.1. – Knowledge bases
- A3.2.2. – Knowledge extraction, cleaning
- A3.2.3. – Inference
- A3.2.4. – Semantic Web
- A3.3. – Data and knowledge analysis
- A3.3.1. – On-line analytical processing
- A3.3.2. – Data mining
- A3.3.3. – Big data analysis
- A3.4.1. – Supervised learning
- A3.4.2. – Unsupervised learning
- A3.4.3. – Reinforcement learning
- A3.4.4. – Optimization and learning
- A3.4.5. – Bayesian methods
- A3.4.6. – Neural networks
- A3.4.8. – Deep learning
- A3.5.2. – Recommendation systems
- A5.1. – Human-Computer Interaction
- A5.2. – Data visualization
- A5.3. – Image processing and analysis
- A5.3.2. – Sparse modeling and image representation
- A5.4.1. – Object recognition
- A5.4.6. – Object localization
- A5.4.7. – Visual servoing
- A9.1. – Knowledge
- A9.2. – Machine learning
- A9.3. – Signal analysis
- A9.4. – Natural language processing

A9.6. – Decision support

A9.7. – AI algorithmics

A9.8. – Reasoning

A9.10. – Hybrid approaches for AI

**Other research topics and application domains**

B3.5. – Agronomy

B3.6. – Ecology

B3.6.1. – Biodiversity

B9.1. – Education

B9.5.6. – Data science

# 1 Team members, visitors, external collaborators

## Research Scientists

- Luis Galarraga Del Prado [INRIA, Researcher]
- Gonzalo Mendez Cobena [INRIA, Chair, from Nov 2024, IIC (2024-2026)]
- Paul Viallard [INRIA, ISFP, from Oct 2024]

## Faculty Members

- Alexandre Termier [Team leader, Univ. Rennes, Professor, HDR]
- Tassadit Bouadi [Univ. Rennes, Associate Professor]
- Peggy Cellier [INSA RENNES, Associate Professor, HDR]
- Sebastien Ferré [Univ. Rennes, Professor, HDR]
- Elisa Fromont [Univ. Rennes, Professor, HDR]
- Romaric Gaudel [Univ. Rennes, Associate Professor, HDR]
- Christine Largouët [L'INSTITUT AGRO, Associate Professor, HDR]
- Véronique Masson [Univ. Rennes, Associate Professor]
- Laurence Rozé [INSA RENNES, Associate Professor]

## Post-Doctoral Fellows

- Aurélien Lamercherie [Univ. Rennes, from Mar 2024]
- Paul Viallard [INRIA, Post-Doctoral Fellow, from Feb 2024 until Sep 2024]

## PhD Students

- Ismail Bachchar [ORANGE LABS, CIFRE, from Feb 2024, with AIstrosight Team]
- Niels Cobat [Univ. Rennes, with Pacap Team]
- Sacha Germain [INRIA, from Nov 2024]
- Elodie Germani [Univ. Rennes, until Sep 2024, with Empenn Team]
- Yasmine Hachani [INRIA, with Sairpico Team]
- Nouha Karaouli [Univ. Rennes, from Dec 2024, with Shadoc Team]
- Gwladys Kelodjou [Univ. Rennes]
- Charbel Kindji [ORANGE LABS, CIFRE]
- Lucie Lepetit [INRIA]
- Dimitri Lerévérénd [INRIA, with WideTeam]
- Youwan Mahé [SIEMENS, CIFRE, from Nov 2024, with EMPENN Team]
- Pierre Maurand [INSA RENNES]
- Manuel Nkegoum [ATERMES, CIFRE, from Nov 2024, with Obelix Team]
- Paul Sevellec [STELLANTIS, CIFRE]
- Isseïnie Sinouvassane [Univ. Rennes]

## Technical Staff

- Louis Bonneau De Beaufort [L'INSTITUT AGRO, Engineer]
- Julianne Guerbette [INRIA, Engineer]
- Frederic Lang [Univ. Rennes, Engineer, from Feb 2024]

## Interns and Apprentices

- Thibault Chanus [ENS RENNES, Intern, from Mar 2024 until May 2024]
- Niels Cobat [Univ. Rennes, Intern, until Jul 2024]
- Romain Eisenschmidt [INRIA, Intern, from May 2024 until Jul 2024]
- Amina Lahlah [INRIA, Intern, from May 2024 until Jul 2024]
- Enzo Ouattara [INSA RENNES, Intern, from May 2024 until Jul 2024]
- Loane Portier [Univ. Rennes, Apprentice, from Sep 2024]
- Loane Portier [INRIA, Intern, from Feb 2024 until May 2024]

## Administrative Assistant

- Gaelle Tworkowski [INRIA]

## External Collaborator

- Tristan Bitard-Feildel [DGA-MI]

## 2 Overall objectives

Data collection is ubiquitous nowadays and it is providing our society with tremendous volumes of knowledge about human, environmental, and industrial activity. This ever-increasing stream of data holds the keys to new discoveries, both in industrial and scientific domains. However, those keys will only be accessible to those who can make sense out of such data. This is, however, a hard problem. It requires a good understanding of the data at hand, proficiency with the available analysis tools and methods, and good deductive skills. All these skills have been grouped under the umbrella term “Data Science” and universities have put a lot of effort in producing professionals in this field. “Data Scientist” is currently an extremely sought-after job, as the demand far exceeds the number of competent professionals. Despite its boom, data science is still mostly a “manual” process: current data analysis tools still require a significant amount of human effort and know-how. This makes data analysis a lengthy and error-prone process. This is true even for data science experts, and current approaches are mostly out of reach of non-specialists.

**The objective of the team LACODAM is to facilitate the process of making sense out of (large) amounts of data.** This can serve the purpose of deriving knowledge and insights for better decision-making. Our approaches are mostly dedicated to provide novel tools to data scientists, that can either perform tasks not addressed by any other tools, or that improve the performance in some area for existing tasks (for instance reducing execution time, improving accuracy or better handling imbalanced data).

## 3 Research program

### 3.1 Introduction

LACODAM is a research team on data science methods and applications, composed of researchers with a background in symbolic AI, data mining, databases, and machine learning. Our research is organized along the three following research axes:

- **Symbolic methods** (Section 3.2) is the first fundamental research axis. It focuses on methods that operate in symbolic domains, that usually take as input discrete data (ex: event logs, transactional data, RDF data) and output symbolic results (ex: patterns, concepts).
- **Interpretable Machine Learning** (Section 3.3) is the other fundamental research axis of the team. It aims at providing interpretable machine learning approaches, mostly by proposing *post-hoc interpretability* for state-of-the-art numerical machine learning methods. *Interpretable by design* machine learning approaches that do not fall into the "Symbolic methods" axis are also studied here.
- **Real world AI** (Section 3.4) deals with the application or adaptation of the methods developed in the aforementioned fundamental axes to real world problems. These works are conducted in collaboration with either industrial or academic partners from other domains. For example, one important application area for the team is digital agriculture with colleagues from Inrae.

## 3.2 Symbolic methods

LACODAM's core symbolic expertise is in methods for exploring efficiently large combinatorial spaces. Such expertise is used in three main research areas:

- Pattern mining, a field of data mining where the goal is to find regularities in data (in an unsupervised way);
- Semantic web, where the goal is to reason over the contents of the Web;
- Skyline queries, where the goal is to find solutions to multiple criteria optimization queries.

In the pattern mining domain, the team is well known for tackling problems where the data and expected patterns have a temporal components. Usually the data considered are timestamped event logs, an ubiquitous type of data nowadays. The patterns extracted can be more or less complex subsequences, but also patterns exhibiting temporal periodicity.

A well-known problem in pattern mining is pattern explosion: due to either underspecified constraints or the combinatorial nature of the search space, pattern mining approaches may produce millions of patterns of mixed interest. The current best approach to limit the number of output patterns is to produce a small size *pattern set*, where the set optimizes some quality criteria. The best pattern set methods so far are based on information theory and rely on the principle of Minimum Description Length (MDL). LACODAM is the leading French team on MDL-based pattern mining, especially for complex patterns. After having integrated Peggy Cellier in 2021, who is the main French expert in MDL-based pattern mining, we integrated in April 2022 Sébastien Ferré, who is also an expert in this area, especially for graph patterns.

The contribution of the team in the Semantic Web domain focuses on different problems related to knowledge graphs (KGs) – usually extracted (semi-)automatically from the Web. These include applications such as mining and reasoning, as well as data management tasks such as provenance and archiving. Reasoning can resort to either symbolic methods such as Horn rules or numeric approaches such as KG embeddings that can be explained via post-hoc explainability modules. The integration of Sébastien Ferré (former SemLIS team leader) further strengthens the Semantic Web axis by extending our expertise on general graph mining, relation extraction, and semantic data exploration.

Skyline queries is a research topic from the database community, and is closely related to multi-criteria optimization. In transactional data, one may want to optimize over several different attributes of equal importance, which means discovering a Pareto Front (the "skyline"). The team has expertise on skyline queries in traditional databases as well as their application to pattern mining (extraction of *skypatterns*). Recently, the team started to tackle the extraction of skyline *groups*, i.e. groups of records that together optimize multiple criteria.

## 3.3 Interpretable ML

Making Machine Learning more interpretable is one of the greatest challenges for the AI community nowadays. LACODAM contributes to the main areas of explainable AI (XAI):



- From a fundamental point of view, the team is trying to deepen the understanding of state-of-the-art post-hoc interpretability approaches (LIME/SHAP), in order to improve these methods or adapt them to novel domains. The team has also started working on the generation of counterfactual explanations. Both lines of work have in common the need for novel notions of neighborhood of points in the model's data space.
- The team is also working on “interpretable-by-design” machine learning methods, where the decision taken can immediately be explained by the (part of) the model that took the decision. Approaches used can as well be deep learning architectures or hybrid numeric/symbolic models relying on pattern mining techniques.
- Last, the team has a special interest in time series data, which arises in many applications but has not yet received enough attention from the interpretability community. We have proposed both post-hoc and “by design” approaches for interpretable ML for time series.

More generally, LACODAM is interested in the study of the interpretability-accuracy trade-off. Our studies may be able to answer questions such as “how much accuracy can a model lose (or perhaps gain) by becoming more interpretable?”. Such a goal requires us to define interpretability in a more principled way—a challenge that has very recently been addressed, not yet overcome.

### 3.4 Real world AI

LACODAM's research work is firmly rooted in applications. On the one hand the data science tools proposed in our fundamental work need to prove their value at solving actual problems. And on the other hand, working with practitioners allows us to understand better their needs and the limitations of existing approaches w.r.t. those needs. This can open new and fruitful (fundamental) research directions.

Our objective, in that axis, is to work on challenging problems with interesting and pertinent partners. We target problems where off-the-shelf data science approaches either cannot be applied or do not give satisfactory results: such problems are the most likely to lead to new and meaningful research in our field. For some problems, collaborative research may not necessarily lead to fundamental breakthroughs, but can still allow making progress in the practitioners' field. We also value such work, which contributes to the discovery of new knowledge and helps industrial partners innovate.

Due to the team expertise in handling temporal data, a lot of our applicative collaborations revolve around the analysis of time series or event logs. Naturally, our work on interpretability is also present in most of our collaborations, as experts want accurate models, but also want to understand the decisions of those models.

The precise application domains are described in more details in the next section (Section 4).

## 4 Application domains

The current period is extremely favorable for teams working in Data Science and Artificial Intelligence, and LACODAM is not the exception. We are eager to see our work applied in real world applications, and have thus an important activity in maintaining strong ties with industrial partners concerned with marketing and energy as well as public partners working on health, agriculture and environment.

### 4.1 Industry

We present below our industrial collaborations. Some are well-established partnerships, while others are more recent collaborations with local industries that wish to reinforce their Data Science R&D with us.

- **Heterogeneous tabular data generation with deep generative models** Tabular data generation is paramount when dealing with privacy-sensitive data and with missing values, which are frequent cases in the real (industrial) world and particularly at Orange. It is also used for data augmentation, a pre-processing step often needed when training data-hungry deep learning models (for example to detect anomalies in networks, study customer profiles, ...). The CIFRE PhD of Charbel Kindji, funded by Orange, is concerned with this application. We study methods to tackle this problem

when the tabular data are heterogeneous (numerical and symbolic) and when new tables should be generated from scratch based on a human prompt.

- **Counterfactual explanations over multivariate time series.** Very complex machine learning models (that are called black-boxes) are often used in critical applications (e.g. self-driving cars). To comply with EU regulations and better understand their systems, many companies, and in particular Stellantis, are interested in developing skills in "explainable AI", a domain which aims at bringing back the human in the decision loop that involves a black box model. The CIFRE PhD of Paul Sevellec, funded by Stellantis, is concerned with this application. We study the particular case of counterfactual explanations on the challenging context of multivariate time-series. This problem is related to the generation of new data that fulfills some human requirements.
- **Local search for multi-armed bandit problems** Multi-armed bandits is the paradigm to design algorithms which simultaneously learn from the data they have collected and act (and therefore collect data) based on what they have learned. While being important for many applications, such algorithms prove to be inefficient when confronted to combinatorial optimization problems. To remove this limit, we are developing bandit algorithms dedicated to combinatorial problems which can be solved through local search. This project is currently supported by a funding from a collaboration between Inria and DGA-AID to foster research subject which interest for both the army or the industry.
- **Analysis and optimization of 3D-printing files through Machine Learning** In the realm of Additive Manufacturing, and more specifically Fused Filament Fabrication 3D printing, print time estimation and optimization plays a pivotal role. The two main approaches for this task are parametric models based on STL input, and analytical models based on G-code. In the context of the PhD of Niels Cobat, we explore the potential of Machine Learning models dedicated to sequences to handle this tasks.
- **Object Detection from Few Multispectral Examples** This project, developed during the thesis of Manuel Nkegoum, aims at providing robust deep-learning-based methods to detect objects in outdoor environments using multispectral images under a low supervision context. The developed methods are expected to learn from few labeled examples at training time and be able to detect scarcely-observed objects in prediction. In case of very few object labels (even no label) being available, the model to be developed should be capable of discovering unknown novel objects from the observed scene.
- **Anomaly detection and segmentation for the characterization of post-stroke recovery.** Stroke is a major health issue globally, causing severe brain damage due to disrupted blood supply. Medical imaging, especially MRI, is crucial for assessing stroke localization and extent. Our goal in this project, with the thesis of Youwan Mahé, is to improve the detection and delineation of chronic stroke lesions from multimodal data using deep learning, helping clinicians plan better treatment and rehabilitation programs.
- **Generation of stable and robust explanations.** This project, funded by Orange, aims to generate robust and reliable local individual explanations, considering data drift when the model's execution data differ from the training data. The goal is to ensure explanations remain valid across different distributions, focusing on mixed tabular data (numerical and categorical). Another promising direction that we identify is how can causality improve current xAI methods, especially in terms of robustness, generalization across domains/tasks, and safety.

## 4.2 Agriculture and Environment

- **Animal welfare.** There has been an increasing concern of both consumers and professionals to better take into account farm animals welfare. For consumers, this is an important ethical issue. For professionals, their animals will have to be able to adapt to quickly evolving climatic conditions due to global warming, thus required to improve animal health and resilience. Better understanding animal welfare in a key component of these improvements. This is the general topic of the WAIT4

project (see Section 10.3), where Lacodam provides its data mining expertise to analyze time series of precision farming sensors, as well as event logs of animal behaviors. As a first topic of research in this project, the PhD of Lucie Lepetit is concerned with heat stress. The data are rumen temperature data from dairy cows of our Inrae partner. In this data, we can notice that in especially hot days of summer, some cows have difficulties to cope with the high temperature and while exhibit high rumen temperature both during the event and during several days after. While on the other hand, there are cows that are only mildly affected by the heat during the event, and who will quickly resume to a normal rumen temperature. Our goal is to design a method that quickly identifies all the abnormal rumen temperature periods correlated to high external temperature, and that provides a characterization of the cows that either resist well to the heat, or on the contrary do not cope well with it. A second topic is to better understand the behavior of animals in “normal” conditions, thanks to the analysis of constant monitoring data. The PhD goal of Sacha Germain, started in november 2024, is to propose methods for identifying individuals’ well-being levels by focusing on both their individual activities and their relationships within the group. The assessment of well-being will rely on behavior analysis, which will be automatically learned from time series data or logs. The approach will aim to develop interpretable models with extend the PhD works of Lénaïg Cornanguer, which defended her PhD in the Lacodam team in 2023.

- **Prediction of the Dynamics of Crop Diseases.** The PhD thesis of Olivier Gauriau (defended on November 2024) focused on the prediction of the dynamics of crop diseases by means of pattern-aided regression techniques. Such techniques are known to strike an interesting trade-off between accuracy and interpretability, which can help agronomers understand the best predictors of high disease incidence, and therefore optimize the usage of phytosanitary products. This project was funded by the #DigitAg convergence lab and the Ecophyto program, and constituted a collaboration with the ACTA of Toulouse and the INRAE.
- **Deep learning-based analysis of the early development of bovine embryos from videomicroscopy.** The PhD of Yasmine Hachani (collaboration with team Sairpico and INRAE) focuses on designing deep learning methods for the comparison and classification of videos of embryos produced in vitro (PIV). These automatic methods are eagerly awaited by biologists in order to broaden the potential of fundamental and applied research in this field, and to help improve results and reproductive performance in breeding. The problem posed is multifaceted. First of all, the images acquired by microscopy are complex in nature: they are low-contrast, noisy, contain transparency effects, and movements are difficult to characterize. The categorization of in vitro fertilized embryos, in terms of the quality of their development, is based on a continuum of classes, rather than distinct ones. Furthermore, the need is to obtain reliable classification at the earliest possible stage, i.e. 3 days post-gamete contact, from a video of 300 images, with images acquired every 15 minutes. Finally, while classification can be supervised, we have only a limited amount of data (a few hundred videos) for deep learning purposes, especially as class characterization can only be achieved by observing a video in its entirety.

### 4.3 Cognitive Sciences

- **Detecting high cognitive load.** Being able to identify whether a particular task incurs a high cognitive load among people is of utter importance in different domains such as education, communication, and design. So far, existing solutions to this problem are either too intrusive (i.e., they require wearable devices with electrodes) or they rely on fully subjective reports. Through the joint collaboration between Miguel Nacenta from University of Victoria, Rodne Quijije from ESPOL (Escuela Superior Politécnica del Litoral in Ecuador) and the LACODAM team (Luis Galárraga and Gonzalo Méndez), we are studying non-intrusive, objective, and low-cost solutions to this problem. Our approach resorts to a secondary repetitive task that consists of drawing circles on a tablet during the execution of the primary task whose cognitive load interests us. Those circular traces can be treated as multivariate time series and their properties can help us elucidate whether the participant is being cognitively challenged or not. The analysis of such time series data resorts to explainable AI techniques, namely SOTA time classifiers and post-hoc explainability techniques. This

is so because understanding the links between high cognitive load and the geometric properties of the traces is crucial to understand how humans behave faced to difficult intellectual tasks.

#### 4.4 Semantic Data Management

- **RDF Archiving and Provenance.** Archiving and provenance tracking are two crucial tasks in the management of large collaborative RDF knowledge bases, such as Wikidata or DBpedia. This is a consequence of the dynamicity and source heterogeneity of such data collections. Notwithstanding the value of RDF archiving and provenance tracking for both data maintainers and consumers, this field of research remains under-developed for multiple reasons. These include, among others, the lack of usability and scalability of the existing systems, a disregard of the evolution patterns of RDF datasets, and a weaker focus on data processes involving non-monotone operations<sup>1</sup>. These challenges are tackled in our ongoing collaboration with the DAISY team of Aalborg University, namely thanks the PhD thesis of Olivier Pelgrin on scalable RDF archiving, and the post-doctoral fellowship of Daniel Hernández on how-provenance computation for SPARQL queries.

## 5 Social and environmental responsibility

### 5.1 Footprint of research activities

There are two main axes that characterize the bulk of LACODAM's environmental impact: work trips, and computing resources utilisation.

**Work trips.** Whenever possible, we prefer using train rather than plane for national and European travels. Most of us continue to submit papers to international conferences outside of Europe but if a paper gets accepted into such conference, we prioritize sending the first author (PhD student). Outside of conferences, for national events (seminars, PhD juries, etc.), videoconference is increasingly used, which helps to reduce the overall carbon footprint of the community.

**Utilisation of computing resources.** The discontinuation of Igrida services and the transition towards Grid'5000 and Jean Zay has reduced our access to easily available computation resources. It adds friction to making experiments, but as a positive effect on energy consumption, as we are now using national infrastructures that benefit from even better sharing between users than Igrida (which was already heavily used).

### 5.2 Impact of research results

We estimate that the research work can have actual impact in three different ways:

- In the short/medium term, a significant part of our research work is conducted in collaboration with companies, through CIFRE PhDs. Hence, the addressed research problems concern an important challenge for the company, and the solutions proposed are evaluated on their relevance to tackle this challenge.
- In the medium/long term, we also have potential impactful research work with scientists from other domains, especially in environment and agriculture. Some earlier work of the team, conducted with INRAE SAS team, helped better understand nitrate pollution in Brittany, an important environmental issue. Current work on the WAIT4 project is dedicated to the design of better data mining tools to characterize heat stress for the cows, which will help to guarantee the well-being of farm animals in a time of climate change.
- Last, in the longer term, the team has a fundamental line of work on machine learning and interpretability. Given the increasing use of machine learning solutions in most areas of human activity, work on interpretability is of utmost societal importance, as it will help in designing more useful

---

<sup>1</sup>Processes where there is some sort of data difference

and also more acceptable machine learning approaches. This will require a sustained effort from the community: LACODAM is taking part in this effort with an important number of contributions this area.

## 6 Highlights of the year

The main highlight of year 2024 was the hiring of Paul Viallard as an INRIA ISFP researcher. This is a permanent position, such hirings are rare: this is a testimony to the academic excellence of Paul, and his brilliant performance in front of the jury.

We are also pleased to welcome another excellent researcher in the team: Gonzalo Mendez Cobena was awarded an Inria International Chair, and will thus make regular visits to our team for the coming 5 years.

Last but not least, Luis Galarraga Del Prado and Christine Largouët were invited on the “L’Esprit Sorcier TV” YouTube channel (follow up of the famous “C’est pas sorcier” TV program) to present AI explainability concepts. The video is at [this link](#). We are very proud of this mediation activity by our colleagues.

### 6.1 Awards

- L’Oréal/Unesco Young Talent Award for Women in Science for Elodie Germani
- “Digital - electronics, photonics, AI and cyber” prize, Trophées Valorisation 2024 for Elisa Fromont
- Best Poster IABM 2024 for Elodie Germani
- Sébastien Ferré was invited at the ARCathon Week, organized by Lab42 in Davos on 4-8 March, after earning a Novelty Prize at the [ARCathon 2023](#) competition.

## 7 New software, platforms, open data

### 7.1 New software

#### 7.1.1 HIPAR

**Name:** Hierarchical Interpretable Pattern-aided Regression

**Keywords:** Regression, Pattern extraction

**Functional Description:** Given a (tabular) dataset with categorical and numerical attributes, HIPAR is a Python library that can extract accurate hybrid rules that offer a trade-off between (a) interpretability, (b) accuracy, and (c) data coverage.

**URL:** <https://gitlab.inria.fr/lgalarra/hipar>

**Contact:** Luis Galarraga Del Prado

#### 7.1.2 Dexteris

**Keywords:** Data Exploration, Querying, Interactive method, JSON

**Functional Description:** Dexteris is a low-code tool for data exploration and transformation. It works as an interactive data-oriented query builder with JSONiq as the target query language. It uses JSON as the pivot data format but it can read from and write to a few other formats: text, CSV, and RDF/Turtle (to be extended to other formats).

Dexteris is very expressive as JSONiq is Turing-complete, and supports a varied set of data processing features: - reading JSON files, and CSV as JSON (one object per row, one field per column), - string processing (split, replace, match, ...), - arithmetics, comparison, and logics, - accessing and

creating JSON data structures, iterations, grouping, filtering, aggregates and ordering (FLWOR operators), - local function definitions.

The built JSONiq programs are high-level, declarative, and concise. Under-progress results are given at every step so that users can keep focused on their data and on the transformations they want to apply.

**URL:** <http://www.irisa.fr/LIS/ferre/dexteris/>

**Publication:** [hal-04186117v1](https://hal.archives-ouvertes.fr/hal-04186117v1)

**Contact:** Sebastien Ferre

### 7.1.3 skm

**Name:** scikit-mine

**Keywords:** Artificial intelligence, Data mining, Pattern discovery, Sequential patterns

**Functional Description:** The library offers several algorithms for extracting a reasonable-sized set of patterns for different types of data (itemsets, sequences, graphs).

**URL:** <https://scikit-mine.github.io/scikit-mine/>

**Contact:** Peggy Cellier

## 8 New results

We organize the scientific results of the research conducted at LACODAM according to the axes described in our research program (Section 3).

### 8.1 Symbolic Methods

#### 8.1.1 Pattern Mining

**Participants:** Tassadit Bouadi, Sébastien Ferré, Luis Galárraga.

*Remark about the “Participants” boxes: we compiled syntactically the list of co-authors of the papers that make the “New Results” of the year, for each subsection. It obviously does not mean that other members of the team do not work on the topics listed, the correct meaning is that they did not have a publication on that topic this year.*

**Neurosymbolic Methods for Rule Mining [39].** In this chapter, we address the problem of rule mining, beginning with essential background information, including measures of rule quality. We then explore various rule mining methodologies, categorized into three groups: inductive logic programming, path sampling and generalization, and linear programming. Following this, we delve into neurosymbolic methods, covering topics such as the integration of deep learning with rules, the use of embeddings for rule learning, and the application of large language models in rule learning.

**Tackling the Abstraction and Reasoning Corpus (ARC) with Object-centric Models and the MDL Principle [26].** The Abstraction and Reasoning Corpus (ARC) is a challenging benchmark, introduced to foster AI research towards human-like intelligence. It is a collection of unique tasks about generating colored grids, specified by a few examples only. In contrast to the transformation-based programs of existing work, we introduce object-centric models that are in line with the natural programs produced by humans. Our models can not only perform predictions, but also provide joint descriptions for input/output pairs. The Minimum Description Length (MDL) principle is used to efficiently search the

large model space. A diverse range of tasks are solved, and the learned models are similar to natural programs.

**SSDBM '24: Proceedings of the 36th International Conference on Scientific and Statistical Database Management [41].** The International Conference on Scientific and Statistical Database Management (SSDBM) brings together scientific domain experts, database researchers, practitioners, and developers for the presentation and exchange of current research results on concepts, tools, and techniques for scientific and statistical database applications. SSDBM 2024 continues the tradition of past SSDBM conferences in providing a stimulating environment to encourage discussion, fellowship, and exchange of ideas in all aspects of research related to scientific and statistical data management. The conference was held from July 10 to 12, 2024, at the Inria Centre at Rennes University, Rennes, France.

### 8.1.2 Graph-FCA

**Participants:** Tassadit Bouadi, Peggy Cellier, Sébastien Ferré, Luis Galárraga.

Some of previously presented documents also contribute to this research domain: [41] .

**Conceptual Knowledge Structures [40].** This book constitutes the proceedings of the First International Joint Conference on Conceptual Knowledge Structures, CONCEPTS 2024, which took place in Cádiz, Spain, during September 9-13, 2024. The conference is an amalgamation of the 18th International Conference on Formal Concept Analysis (ICFCA); the 17th International Conference on Concept Lattices and Their Applications (CLA); and the 28th International Conference on Conceptual Structures (ICCS). The 18 full and 4 short papers included in this book were carefully reviewed and selected from 38 submissions. They were organized in topical sections as follows: Theory; algorithms, methods, and resources; applications.

**Comparing Relational Concept Analysis and Graph-FCA on their Common Ground [35].** Relational Concept Analysis (RCA) and Graph-FCA (GCA) have been defined as Formal Concept Analysis (FCA) extensions for processing relational data and knowledge graphs respectively. Nevertheless, while their purposes and results seem similar, the data modelling and the definition of concepts are different. In this contribution, we compare these two approaches on a common basis, considering only unary and binary relations for GCA and the existential quantifier for RCA. We focus on examples showing the similarities and dissimilarities between both methods, and highlighting how cycles are processed differently by RCA and GCA

### 8.1.3 Semantic Web

**Participants:** Tassadit Bouadi, Peggy Cellier, Sébastien Ferré, Luis Galárraga.

Some of previously presented documents also contribute to this research domain: [41] .

**NPCS: Native Provenance Computation for SPARQL [25].** <div><p>The popularity of Knowledge Graphs (KGs) both in industry and academia owes credit to their flexible data model, suitable for data integration from multiple sources. Several KG-based applications such as trust assessment or view maintenance on dynamic data rely on the ability to compute provenance explanations for query results. The how-provenance of a query result is an expression that encodes the records (triples or facts) that explain its inclusion in the result set. This contribution proposes NPCS, a Native Provenance Computation approach for SPARQL queries. NPCS annotates query results with their how-provenance. By building upon spm-provenance semirings, NPCS supports both monotonic and non-monotonic SPARQL queries. Thanks to its reliance on query rewriting techniques, the approach is directly applicable to already deployed SPARQL engines using different reification schemes -including RDF-star. Our experimental

evaluation on two popular SPARQL engines (GraphDB and Stardog) shows that our novel query rewriting brings a significant runtime improvement over existing query rewriting solutions, scaling to RDF graphs with billions of triples.

### **Investigating the use of language models for natural language querying over knowledge graphs [31].**

In this contribution, we present the results of an in-depth study on the performance of large language models (LLMs) in the context of natural language knowledge graph querying (KGQA). The experimental methodology was structured around two distinct approaches: SPARQL query generation and direct querying. The results on the QALD-10 benchmark revealed very poor performance in the first approach and satisfactory performance in the second, with significant variations depending on the type of questions and answers.

## **8.2 Interpretable Machine Learning**

**Participants:** Julien Delaunay, Luis Galárraga, Romaric Gaudel, Gwladys Kelodjou, Christine Largouët, Véronique Masson, Laurence Rozé, Alexandre Terrier.

**A PAC-Bayesian Bound on a Risk Measure for Fair Learning [37].** We study the Conditional Value at Risk (CVaR), a risk measure defined as a combination of subgroup risks that focuses solely on the highest risks. In fair learning, a subgroup is defined by the set of individuals sharing the same value of the sensitive attribute under consideration. CVaR is thus valuable as it concentrates on the values of the sensitive attribute associated with high risk. In this work, we derive a PAC-Bayesian generalization bound for CVaR that involves not only a distribution over the set of hypotheses (as is common in PAC-Bayes), but also a distribution over the set of sensitive attribute values, allowing control over the importance of each subgroup contributing to the CVaR.

**Explaining a Black Box Without a Black Box [21].** Counterfactual explanation methods are popular approaches for explaining machine learning algorithms. These explanations encode the necessary modifications in a target document to change a classifier's prediction. Most of these methods find such explanations by iteratively perturbing the target document until it is classified differently by the black box. We identify two main families of counterfactual approaches in the literature: (a) "transparent" methods that perturb the target by adding, removing, or replacing words, and (b) "opaque" techniques that project the target document into a non-interpretable latent space where the perturbation is then performed. This contribution presents a comparative study of the performance of these two families of methods on three standard natural language processing tasks. Our results show that for applications such as fake news detection or sentiment analysis, opaque counterfactual approaches may introduce an additional level of complexity without significant improvement.

**Uniform Generalization Bounds on Data-Dependent Hypothesis Sets via PAC-Bayesian Theory on Random Sets [43].** We propose data-dependent uniform generalization bounds by approaching the problem from a PAC-Bayesian perspective. We first apply the PAC-Bayesian framework on 'random sets' in a rigorous way, where the training algorithm is assumed to output a data-dependent hypothesis set after observing the training data. This approach allows us to prove data-dependent bounds, which can be applicable in numerous contexts. To highlight the power of our approach, we consider two main applications. First, we propose a PAC-Bayesian formulation of the recently developed fractal-dimension-based generalization bounds. The derived results are shown to be tighter and they unify the existing results around one simple proof technique. Second, we prove uniform bounds over the trajectories of continuous Langevin dynamics and stochastic gradient Langevin dynamics. These results provide novel information about the generalization properties of noisy algorithms.



**A PAC-Bayesian Link Between Generalisation and Flat Minima [47].** Modern machine learning usually involves predictors in the overparametrised setting (number of trained parameters greater than dataset size), and their training yield not only good performances on training data, but also good generalisation capacity. This phenomenon challenges many theoretical results, and remains an open problem. To reach a better understanding, we provide novel generalisation bounds involving gradient terms. To do so, we combine the PAC-Bayes toolbox with Poincaré and Log-Sobolev inequalities, avoiding an explicit dependency on dimension of the predictor space. Our results highlight the positive influence of flat minima (being minima with a neighbourhood nearly minimising the learning problem as well) on generalisation performances, involving directly the benefits of the optimisation phase.

**Shaping Up SHAP: Enhancing Stability through Layer-Wise Neighbor Selection [29].** Machine learning techniques, such as deep learning and ensemble methods, are widely used in various domains due to their ability to handle complex real-world tasks. However, their black-box nature has raised multiple concerns about the fairness, trustworthiness, and transparency of computer-assisted decision-making. This has led to the emergence of local post-hoc explainability methods, which offer explanations for individual decisions made by black-box algorithms. Among these methods, Kernel SHAP is widely used due to its model-agnostic nature and its well-founded theoretical framework. Despite these strengths, Kernel SHAP suffers from high instability: different executions of the method with the same inputs can lead to significantly different explanations, which diminishes the utility of post-hoc explainability. The contribution of this work is twofold. On the one hand, we show that Kernel SHAP's instability is caused by its stochastic neighbor selection procedure, which we adapt to achieve full stability without compromising explanation fidelity. On the other hand, we show that by restricting the neighbors generation to perturbations of size 1—which we call the coalitions of Layer 1—we obtain a novel feature attribution method that is fully stable, computationally efficient, and still meaningful.

**Synthetic Data: Generate Avatar Data on Demand [30].** Anonymization is crucial for the sharing of personal data in a privacy-aware manner yet it is a complex task that requires to set up a trade-off between the robustness of anonymization (i.e., the privacy level provided) and the quality of the analysis that can be expected from anonymized data (i.e., the resulting utility). Synthetic data has emerged as a promising solution to overcome the limits of classical anonymization methods while achieving similar statistical properties to the original data. Avatar-based approaches are a specific type of synthetic data generation that rely on local stochastic simulation modeling to generate an avatar for each original record. While these approaches have been used in healthcare, their attack surface is not well documented and understood. In this contribution, we provide an extensive assessment of such approaches and comparing them against other data synthesis methods. We also propose an improvement based on conditional sampling in the latent space, which allows synthetic data to be generated on demand (i.e., of arbitrary size). Our empirical analysis shows that avatar-generated data are subject to the same utility and privacy trade-off as other data synthesis methods with a privacy risk more important on the edge data, which correspond to records that have the fewest alter egos in the original data.

**A Theoretically Grounded Extension of Universal Attacks from the Attacker's Viewpoint [32].** We extend universal attacks by jointly learning a set of perturbations to choose from to maximize the chance of attacking deep neural network models. Specifically, we embrace the attacker's perspective and introduce a theoretical bound quantifying how much the universal perturbations are able to fool a given model on unseen examples. An extension to assert the transferability of universal attacks is also provided. To learn such perturbations, we devise an algorithmic solution with convergence guarantees under Lipschitz continuity assumptions. Moreover, we demonstrate how it can improve the performance of state-of-the-art gradient-based universal perturbation. As evidenced by our experiments, these novel universal perturbations result in more interpretable, diverse, and transferable attacks.

**Tighter Generalisation Bounds via Interpolation [49].** This contribution contains a recipe for deriving new PAC-Bayes generalisation bounds based on the  $(f, \Gamma)$ -divergence, and, in addition, presents PAC-Bayes generalisation bounds where we interpolate between a series of probability divergences (including but not limited to KL, Wasserstein, and total variation), making the best out of many worlds depending

on the posterior distributions properties. We explore the tightness of these bounds and connect them to earlier results from statistical learning, which are specific cases. We also instantiate our bounds as training objectives, yielding non-trivial guarantees and practical performances.

**Leveraging PAC-Bayes Theory and Gibbs Distributions for Generalization Bounds with Complexity Measures [34].** In statistical learning theory, a generalization bound usually involves a complexity measure imposed by the considered theoretical framework. This limits the scope of such bounds, as other forms of capacity measures or regularizations are used in algorithms. In this contribution, we leverage the framework of disintegrated PAC-Bayes bounds to derive a general generalization bound instantiable with arbitrary complexity measures. One trick to prove such a result involves considering a commonly used family of distributions: the Gibbs distributions. Our bound stands in probability jointly over the hypothesis and the learning sample, which allows the complexity to be adapted to the generalization gap as it can be customized to fit both the hypothesis class and the task.

## 8.3 Real World AI

### 8.3.1 Computer Vision and Robotics

**Participants:** Élisabeth Fromont, Elodie Germani, Yasmine Hachani.

**Uncovering communities of pipelines in the task-fMRI analytical space [27].** Analytical workflows in functional magnetic resonance imaging are highly flexible with limited best practices as to how to choose a pipeline. While it has been shown that the use of different pipelines might lead to different results, there is still a lack of understanding of the factors that drive these differences and of the stability of these differences across contexts. We use community detection algorithms to explore the pipeline space and assess the stability of pipeline relationships across different contexts. We show that there are subsets of pipelines that give similar results, especially those sharing specific parameters (e.g. number of motion regressors, software packages, etc.). Those pipeline-to-pipeline patterns are stable across groups of participants but not across different tasks. By visualizing the differences between communities, we show that the pipeline space is mainly driven by the size of the activation area in the brain and the scale of statistic values in statistic maps.

**The HCP multi-pipeline dataset: an opportunity to investigate analytical variability in fMRI data analysis [45].** Results of functional Magnetic Resonance Imaging (fMRI) studies can be impacted by many sources of variability including differences due to: the sampling of the participants, differences in acquisition protocols and material but also due to different analytical choices in the processing of the fMRI data. While variability across participants or across acquisition instruments have been extensively studied in the neuroimaging literature the root causes of analytical variability remain an open question. Here, we share the *HCP multi-pipeline dataset*, including the resulting statistic maps for 24 typical fMRI pipelines on 1,080 participants of the HCP-Young Adults dataset. We share both individual and group results - for 1,000 groups of 50 participants - over 5 motor contrasts. We hope that this large dataset covering a wide range of analysis conditions will provide new opportunities to study analytical variability in fMRI.

**Predicting Parkinson's disease trajectory using clinical and functional MRI features: a reproduction and replication study [44].** Parkinson's disease (PD) is a common neurodegenerative disorder with a poorly understood pathophysiology and no established biomarkers for the diagnosis of early stages and for prediction of disease progression. Several neuroimaging biomarkers have been studied recently, but these are susceptible to several sources of variability related for instance to cohort selection or image analysis. In this context, an evaluation of the robustness of such biomarkers to variations in the data processing workflow is essential. This study is part of a larger project investigating the replicability of

potential neuroimaging biomarkers of PD. Here, we attempt to reproduce (re-implementing the experiments with the same data, same method) and replicate (different data and/or method) the models described in [52] to predict individual's PD current state and progression using demographic, clinical and neuroimaging features (fALFF and ReHo extracted from resting-state fMRI). We use the Parkinson's Progression Markers Initiative dataset (PPMI, ppmi-info.org), as in [52] and aim to reproduce the original cohort, imaging features and machine learning models as closely as possible using the information available in the paper and the code. We also investigated methodological variations in cohort selection, feature extraction pipelines and sets of input features. Different criteria were used to evaluate the reproduction and compare the reproduced results with the original ones. Notably, we obtained significantly better than chance performance using the analysis pipeline closest to that in the original study ( $R^2 \geq 0$ ), which is consistent with its findings. Moreover, using derived data provided by the authors of the original study, we were able to make an exact reproduction and managed to obtain results that were close to the original ones. The challenges encountered while reproducing and replicating the original work are likely explained by the complexity of neuroimaging studies, in particular in clinical settings. We provide recommendations to further facilitate the reproducibility of such studies in the future.

**Mitigating analytical variability in fMRI results with style transfer [46].** We propose a novel approach to improve the reproducibility of neuroimaging results by converting statistic maps across different functional MRI pipelines. We make the assumption that pipelines used to compute fMRI statistic maps can be considered as a style component and we propose to use different generative models, among which, Generative Adversarial Networks (GAN) and Diffusion Models (DM) to convert statistic maps across different pipelines. We explore the performance of multiple GAN frameworks, and design a new DM framework for unsupervised multi-domain style transfer. We constrain the generation of 3D fMRI statistic maps using the latent space of an auxiliary classifier that distinguishes statistic maps from different pipelines and extend traditional sampling techniques used in DM to improve the transition performance. Our experiments demonstrate that our proposed methods are successful: pipelines can indeed be transferred as a style component, providing an important source of data augmentation for future medical studies.

**Early prediction of the transferability of bovine embryos from videomicroscopy [50], [28], [36].** We aim to predict early and automatically the transferability of embryos for biological issues and cattle breeding challenges, taking as input 2D time-lapse videos up to the fourth embryonic development day. We propose: 1- a formulation as a supervised classification with two classes Transferable (T) and Non Transferable (NT), 2- a 3D-CNN with three pathways, multi-scale in time and able to handle appearance and motion in different ways, 3- the use of the focal loss for training. Our model SFR compares favorably to other methods and demonstrates its accuracy and efficiency for our challenging biological task.

### 8.3.2 Agriculture

**Participants:** Élixa Fromont, Luis Galárraga, Olivier Gauriau, Yasmine Hachani, Alexandre Termier.

Some of previously presented documents also contribute to this research domain: [50] .

**Comparing machine-learning models of different levels of complexity for crop protection: A look into the complexity-accuracy tradeoff [22].** Crop diseases and pests constitute significant causes of yield losses for crops. To limit the harm incurred by those events, farmers resort to plant protection products. Such products are known to have adverse effects both on the environment and on human health. Agronomists make continuous efforts to limit the usage of plant protection products to situations where those products are strictly necessary. To determine such situations, agronomists and policy-makers often rely on decision support tools to model and predict the dynamics of plant diseases. Decision support tools are based either on mechanistic models or on statistical approaches learned from large datasets of biotic (e.g., disease incidence, plant phenological stage) and abiotic (meteorological, soil characteristics)

observations in cultures. The surge of powerful machine learning (ML) methods in the last decade makes such approaches a natural pathway to model the dynamics of plant diseases. Machine learning models can reveal the factors that contribute the most to disease and pests outbreaks, provided that those models are simple enough for human inspection. Simplicity, however, may come at the price of lower prediction performances when compared to more complex models. In this contribution, we offer a deep look at the performance of ML models of different complexity when used on two use cases of crop disease prediction: downy mildew in the grapevine, and *Cercospora* leaf spot in the sugar beet. We compare model accuracy and complexity using a year-based cross-validation approach. Our results suggest that interannual meteorological variations are a very important factor in plant disease prediction. Moreover, in line with the observations of the research community in interpretable ML, model complexity stands in clear trade-off with accuracy. This makes models of intermediate complexity appealing for predicting the dynamics of crop diseases as they can provide explicit insights about the rationale of their predictions.

### 8.3.3 Machine Learning on Sequences

**Participants:** Éliisa Fromont, Luis Galárraga, Antonin Voyez.

#### **Analyzing and explaining privacy risks on time series data: ongoing work and challenges [20].**

Currently, privacy risks assessment is mainly performed as audits conducted by data privacy analysts. In the TAILOR project, we promote a more systematic and automatic approach based on interpretable metrics and formal methods to evaluate privacy risks and to control the tension between data privacy and utility. In this contribution, we focus on privacy risks raised by publishing time series datasets, and we survey the methods developed in TAILOR to analyse and quantify privacy risks depending on different publisher and attacker models.

**Fast and Accurate Context-Aware Basic Block Timing Prediction using Transformers [24].** This contribution introduces ORXESTRA, a context-aware execution time prediction model based on Transformers XL, specifically designed to accurately estimate performance in embedded system applications. Unlike traditional machine learning models that often overlook contextual information, resulting in biased predictions for individual isolated basic blocks, ORXESTRA overcomes this limitation by incorporating execution context awareness. By doing so, ORXESTRA effectively accounts for the processor micro-architecture without explicitly modeling micro-architectural elements such as caches, pipelines, and branch predictors. Our evaluations demonstrate ORXESTRA's ability to provide precise timing estimations for different ARM targets (Cortex M4, M7, A53, and A72), surpassing existing machine learning-based approaches in both prediction accuracy and prediction speed.

**REFINE-LM: Mitigating Language Model Stereotypes via Reinforcement Learning [33].** With the introduction of (large) language models, there has been significant concern about the unintended bias such models may inherit from their training data. A number of studies have shown that such models propagate gender stereotypes, as well as geographical and racial bias, among other biases. While existing works tackle this issue by preprocessing data and debiasing embeddings, the proposed methods require a lot of computational resources and annotation effort while being limited to certain types of biases. To address these issues, we introduce REFINE-LM, a debiasing method that uses reinforcement learning to handle different types of biases without any fine-tuning. By training a simple model on top of the word probability distribution of a LM, our bias agnostic reinforcement learning method enables model debiasing without human annotations or significant computational resources. Experiments conducted on a wide range of models, including several LMs, show that our method (i) significantly reduces stereotypical biases while preserving LMs performance; (ii) is applicable to different types of biases, generalizing across contexts such as gender, ethnicity, religion, and nationality-based biases; and (iii) it is not expensive to train.

## 9 Bilateral contracts and grants with industry

### 9.1 Bilateral contracts with industry

- **Stellantis - Univ. Rennes**

**Participants:** Elisa Fromont, Romaric Gaudel, Laurence Rozé, Paul Sévellec.

Contract amount: 70k€ + Phd Salary

Context. This project is a collaboration with Stellantis and focuses on the development of interpretable machine learning models for multivariate time series data. Utilizing a range of sensors integrated within vehicles, these models are designed to make real-time decisions. Providing drivers with clear explanations of these decisions is a key aspect. We specifically concentrate on counterfactual explanations, which not only clarify why a particular decision was made but also illustrate how alternative scenarios might have led to different outcomes.

Objective. Current approaches providing counterfactual explanations for time series models are limited to univariate time series. In this project, we aim to develop approaches to handle multivariate time series, which requires capturing the correlations between the series.

*Additional remarks.* This is the doctoral contract for the PhD of Paul Sévellec (Thèse CIFRE).

- **ORANGE - Univ. Rennes**

**Participants:** Elisa Fromont, Charbel Kindji.

Contract amount: 45k€ + Phd Salary

Context. Tabular data generation is paramount when dealing with privacy-sensitive data and with missing values, which are frequent cases in the real (industrial) world and particularly at Orange. It is also used for data augmentation, a pre-processing step often needed when training data-hungry deep learning models (for example to detect anomalies in networks, study customer profiles, ...).

Objective. We study methods to tackle heterogeneous tabular data generation with deep generative models. We are particularly interested in problems where the tabular data are heterogeneous (numerical and symbolic) and when new tables should be generated from scratch based on a human prompt.

*Additional remarks.* This is the doctoral contract for the PhD of Charbel Kindji (Thèse CIFRE).

- **ATERMES - Univ. Rennes**

**Participants:** Elisa Fromont, Manuel Nkegoum.

Contract amount: 0€ (for LACODAM Team) + Phd Salary

This project aims at providing robust deep-learning-based methods to detect objects in outdoor environments using multispectral images under a low supervision context. The developed methods are expected to learn from few labeled examples at training time and be able to detect scarcely-observed objects in prediction. In case of very few object labels (even no label) being available, the

model to be developed should be capable of discovering unknown novel objects from the observed scene.

*Additional remarks.* This is the CIFRE PhD of Manuel Nkegoum with Atermes (Thèse CIFRE). There is an agreement with the Obelix team to freely use part of the 60k€ contract as was done conversely in the previous PhD with the same parties.

- **SIEMENS - Univ. Rennes**

**Participants:** Elisa Fromont, Youwan Mahé.

Contract amount: 12k€ (for LACODAM Team)+ Phd Salary

Stroke is a major health issue globally, causing severe brain damage due to disrupted blood supply. Medical imaging, especially MRI, is crucial for assessing stroke localization and extent.

Our goal in this project, is to improve the detection and delineation of chronic stroke lesions from multimodal data using deep learning, helping clinicians plan better treatment and rehabilitation programs.

*Additional remarks.* This is the CIFRE PhD of Youwan Mahé with Siemens (Thèse CIFRE). The total contract with Siemens is 50k€ but this amount is divided between the CHU of Rennes, the Empenn team and the Lacodam team.

- **Orange - Univ. Rennes - Univ. Lyon1 - Inria Lyon (AIstroSight)**

**Participants:** Tassadit Bouadi, Ismail Bachchar.

Contract amount: 10k€ (for LACODAM Team) + Phd Salary

Context. This project is a collaboration with Orange Labs Lannion about interpretable machine learning. The Orange company aims to develop the use of machine learning algorithms to enhance the services they propose to their customers (for instance, credit acceptance or attribution prediction). It ensues the development of generic approaches for providing interpretable decisions to customers or client managers.

Orange is committed to the responsible use of AI, emphasizing the importance of algorithm explainability for trustworthy AI and its adoption by individuals. The CIFRE PhD of Isamil Bachchar, funded by Orange, aims to generate robust and reliable local individual explanations, considering data drift when the model's execution data differ from the training data. The goal is to ensure explanations remain valid across different distributions, focusing on mixed tabular data (numerical and categorical).

*Additional remarks.* This contract finances the PhD of Ismail Bachchar by Orange.

## 10 Partnerships and cooperations

### 10.1 International research visitors

#### 10.1.1 Visits of international scientists

**Inria International Chair.** From 2024, and until 2027, LACODAM counts on the expertise of Gonzalo Méndez, a researcher from University of Valencia. Gonzalo is holder of an Inria International Chair and has been a collaborator of the team since 2019. His research work falls within the domain of DataVis

(data visualization) applied to different application settings, including learning analytics and eXplainable AI. Previous work with the team includes the design and evaluation of interactive and visual AI-based systems for course recommendation. As official part of LACODAM Gonzalo will spend in between 6 and 9 months at Inria where he will work with us on two areas in particular:

- Continuing the line of research of the FAbLe project, Gonzalo is working with Luis Galárraga and Christine Largouët in the design and study of narrative-based explanations for AI systems. While the emergence of LLMs has facilitated the automatic use of textual narratives for explainability, our team focuses on scrollytelling explanations: a particular type of narrative that combines text with illustrations as users scroll in the screen. To the best of our knowledge no approach so far has studied the use of scrollytelling for eXplainable AI.
- Gonzalo is also involved in the study of novel methods to predict the cognitive load incurred by users when executing different intellectual tasks. The propose approach analyzes the traces of a repetitive secondary drawing task to infer cognitive effort among people. This project is a collaboration with the University of Victoria and ESPOL (Ecuador) and Luis Galárraga. This project combines expertise from different domains including cognitive sciences, data visualization, and eXplainable AI.

## 10.2 European initiatives

### 10.2.1 Horizon Europe

Several unsuccessful submissions at the European level :

- Elisa Fromont participated in the submission of the PROSPER (Pioneering Responsible disclosure of multimodal healthcare data Openly by data Synthesis and the Prototyping of applications for hearing and balance Evaluation and Restoration) project: call HORIZON-JU-IHI-2023-05 (resubmission in 2025)
- Romaric Gaudel participated in the submission of the DoCA-AI (Intelligent on-demand Detection and Simulation of False Data Injection Attacks in Global Maritime Surveillance) project. call : European Defence Fund-2023-LS-RA-SMERO.  
This project aims to explore novel on-demand interpretable cyber-threat detection and simulation capabilities dedicated to the of the global maritime surveillance and awareness (global MSA). The project focuses on detecting False AIS tracks by leveraging explainable AI methods and combining multi-sources of data.  
The main contribution of LACODAM's members is on the developpment of XAI models and methods relevant dedicated to time-series and therefore relevant for AIS data.

### 10.2.2 H2020 projects

Luis Galarraga Del Prado, Alexandre Termier and Elisa Fromont were part of the **TAILOR** Network within Inria. TAILOR was one of four “Network of Excellence” working on aspects of trustworthy AI funded under the H2020-ICT-48-2020 call.

## 10.3 National initiatives

- **#DigitAg: Digital Agriculture**

**Participants:** Alexandre Termier, Véronique Masson, Christine Largouët, Luis Galárraga, Olivier Gauriau.

#DigitAg is a “Convergence Institute” dedicated to the increasing importance of digital techniques in agriculture. Its goal is twofold: First, making innovative research on the use of digital techniques

in agriculture in order to improve competitiveness, preserve the environment, and offer correct living conditions to farmers. Second, preparing future farmers and agricultural policy makers to successfully exploit such technologies. While #DigitAg is based on Montpellier, Rennes is a satellite of the institute focused on cattle farming.

LACODAM is involved in the “data mining” challenge of the institute, which A. Termier co-leads. He is also the representative of Inria in the steering committee of the institute. The interest for the team is to design novel methods to analyze and represent agricultural data, which are challenging because they are both heterogeneous and multi-scale (both spatial and temporal).

- **PEPR WAIT 4**

**Participants:** Alexandre Termier, Peggy Cellier, Lucie Lepetit, Christine Largouet, Véronique Masson, Louis Bonneau De Beaufort.

The WAIT 4 project is a part of the “Agroecology and numeric” PEPR. The goal of this project is to provide the scientific basis for significant improvements in the well-being of farm animals. Up to now, animal well-being is evaluated with indicators of the means deployed (e.g. available space, method to control building temperature, time spent outside...). The goal of WAIT4 is to provide tools required in order to move to *results* indicators: can some guarantees be given on the well being of animals? Can this well (or unwell) being be correlated to management actions from the farmer, or to their general living conditions?

This requires a much finer understanding of animal mental as well as physiological state. The project is led by Inrae (Florence Gondret), which brings animal science specialists, ranging from biologists to ethologists. CEA provides expertise on blood sensors, to measure molecules linked to stress. And Inria as well as Insa Lyon provide computer science expertise for tools to analyse the data. More precisely, the Lacodam team deals first with analyzing time series of numerical sensor data (e.g. temperature, activity), and second with categorical sequences of events produced by annotation tools from the analysis of videos. Both will help to better model animal behavior, and determine what are “normal” behaviors, and what are anomalous behaviors that may be linked to bad conditions for the animals.

- **Bourse IUF - Elisa FROMONT**

This project supports (until the end of 2024) the work of Elisa Fromont both with a reduction of teaching load, and some research money (15Keuros / year for 5 years). Elisa is currently working on designing effective data mining and machine learning algorithms for real-life data (which are scarce, heterogenous, multimodal, imbalanced, temporal, ...). For the next few years, Elisa would like to focus on the interpretability of the results obtained by these algorithms. In pattern mining, her goal is to design algorithms which can directly mine a small number of relevant patterns. In the case of black box machine learning models (e.g. deep neural nets), Elisa would like to design methods to help the end user understand the decisions taken by the model.

- **PEPR IA ADAPTING**

**Participants:** Elisa Fromont, Nouha Karaouli, Loane Portier.

AdaptING explores new models, computing paradigms (i.e., beyond the Von Neumann architecture), hybrid architectures (i.e., beyond MPSoC – System-on-Chip), and emerging technologies through various initiatives aimed at making AI more efficient, sustainable, and trustworthy. While the project encompasses hardware advancements, our contributions in LACODAM will focus on the algorithmic level. In particular, we will design new resource-efficient incremental learning algorithms that can run on embedded systems with their associated resource and privacy constraints.



We will also investigate post-hoc explanation methods for federated learning systems as a way to monitor the trustworthiness of such systems. Federated learning will often be at the center of the project as a practical learning paradigm suited for embedded systems.

- **Scikit-mine (F-WIN project of PNR-IA)**

**Participants:** Peggy Cellier, Alexandre Termier.

Scikit-mine (SKM for short) is a Python library of pattern mining algorithms, designed to be compatible with the well-known scikit-learn library. It allows practitioners to use state-of-the-art pattern mining algorithms with a library that has the same usage interface as scikit-learn, and that exploits the same data types. SKM was developed by CNRS AI engineers in the context of the F-WIN project of the PNR-IA program of CNRS, whose general goal is to improve the development of AI software in research teams of CNRS labs.

- **Local search for multi-armed bandit problems**

**Participants:** Romaric Gaudel, Elisa Fromont, Paul Viillard.

**Period:** 01/02/2024 – 31/01/2026

**Budget:** 130k€ (Univ Rennes)

This project aims at proposing multi-armed bandit algorithms dedicated to combinatorial problems which can be solved through local search. It is funded through a collaboration between Inria and DGA-AID to foster research subjects which interest either the army or the industry. The fund mainly covers a 2-year postdoc position.

### 10.3.1 ANR

- **FABLe: Framework for Automatic Interpretability in Machine Learning**

Participants: L. Galárraga (holder), C. Largouët

**Participants:** Luis Galárraga (holder), Christine Largouët, Julien Delaunay, Julianne Guerbet.

*How can we fully automatically choose the best explanation for a given use case in classification?* Answering this question is the raison d'être of the JCJC ANR project FABLe. By "best explanation" we mean an explanation that is both understandable by humans and faithful among a universe of possible explanations. We focus on local explanations, i.e., when we want to explain the answer of a black-box model for a given use case, which we call the "target instance". We argue that the choice of the best explanation depends on the (i) data, namely the model, the explanation technique and the target instance, etc., and (ii) the recipients of the explanations. Hence our research is focused on two main questions: "What makes an explanation suitable (interpretable and faithful) for a particular instance and model?" and "What is the effect of the different AI-based explanation techniques and visual representations on users' comprehension and trust?". Answering these questions will help us understand and automate the selection of a particular explanation style based on the use case. Our ultimate goal is to produce a suite of algorithms that will compute suitable explanations for ML algorithms based on our insights of what is interpretable. User studies on different explanation settings (methods and visual representations) will allow us to characterize the features of explanations that make them acceptable (i.e., understandable and trustworthy) by users.

- **SmartFCA: A Smart Tool for Analyzing Complex Data with Formal Concept Analysis**

**Participants:** Sébastien Ferré, Peggy Cellier, Frederic Lang.

**Period:** 01/01/2022 – 31/12/2025

**Budget:** 143k€ (Univ Rennes)

Formal Concept Analysis (FCA) is a mathematical framework based on lattice theory and aimed at data analysis and classification. FCA, which is closely related to pattern mining in knowledge discovery (KD), can be used for data mining purposes in many application domains, e.g. life sciences and linked data. Moreover, FCA is human-centered and provides means for visualization and interaction with data and patterns. Actually it is now possible to deal with complex data such as intervals, sequences, trajectories, trees, and graphs. Research in FCA is dynamic, but there is still room for extensions of the original formalism. Many theoretical and practical challenges remain. Actually there does not exist any consensual platform offering the necessary components for analyzing real-life data. This is precisely the objective of the SmartFCA project to develop the theory and practice of FCA and its extensions, to make the related components inter-operable, and to implement a usable and consensual platform offering the necessary services and workflows for KD.

In particular, for satisfying in the best way the needs of experts in many application domains, SmartFCA will offer a “Knowledge as a Service” (KaaS) component for making domain knowledge operable and reusable on demand.

- **MeKaNo: Search the Web with Things**

**Participants:** Sébastien Ferré, Peggy Cellier, Luis Galárraga, Aurélien Lamer-  
cerie.

**Period:** 01/10/2022 – 29/09/2026

**Budget:** 143k€ (Univ Rennes)

In MeKaNo, we aim to search the web with things, in order to get more accurate results over a wide diversity of sources. Traditional web search engines search the web with strings. However, keyword search often returns many irrelevant documents, pushing users to refine their keyword list following a trial-and-error process. To overcome such limitations, major companies allowed searching for things, not strings. Asking for the age of “James Cameron” to your vocal assistant, it locates in a Knowledge Graph (KG) a Person matching “James Cameron” where a property “age” is set to 66 years, i.e. the Thing “James Cameron”. If searching for Things is a tremendous progress and delivers exact answers, the search is done over a Knowledge Graph and not on the Web. Consequently, there may exist many answers on the web that are not part of the knowledge graph.

To summarize, searching with strings over the web offers diversity at the expense of noise. Searching for Things delivers exact answers, but we lose diversity. In MeKaNo, we aim at searching the web with Things to get diversity and avoid noisy results. To search the web with Things, we face three main scientific challenges:

1. Users are used to search with keywords. Transforming a keyword query into a mixed query that first searches over a KG then into the web is difficult, especially, for complex queries.
2. As with traditional web searches, users expect to obtain ranked results in a snap. Combining KG search and Web search while preserving performances is highly challenging and requires a new kind of search engine.
3. Improving the connection between the web of microdata and Knowledge Graphs requires entity matching at large scale for microdata entities and KG entities.

## 11 Dissemination

### 11.1 Promoting scientific activities

#### 11.1.1 Scientific events: organisation

##### General chair, scientific chair

- Peggy Cellier is member of the steering committee of the European Conference in Machine Learning and Knowledge Discovery (ECML PKDD) since 2022, and until the end of 2025.

##### Member of the organizing committees

- Peggy Cellier is member of the "comité de pilotage" *collège TLH (Technologies du Langage Humain)* of AFIA (Association française pour l'Intelligence Artificielle) since the end of 2024.

#### 11.1.2 Scientific events: selection

##### Chair of conference program committees

- Tassadit Bouadi and Luis Galárraga were the co-organizers of the 7th edition of the *AIMLAI* (Advances in Interpretable ML and AI) workshop that took place in Sept 2024 at the ECML/PKDD conference (in Vilnius). This year workshop featured a tutorial on explainability for sequential and large language models.
- Luis Galárraga was the chair of the PhD Consortium of the IDA (Intelligent Data Analysis) conference that took place in April 2024 in Stockholm.
- Tassadit Bouadi and Luis Galárraga were members of the program and scientific committee of the special session on *Trust in AI: Beyond Interpretability* at the *ESSAN symposium* on October 2024
- Sebastien Ferré was a program chair of the *1st International Joint Conference on Conceptual Knowledge Structures (CONCEPTS)* in September 2024, Cadix. This new conference merges three existing international conferences: ICCS, ICFCA, and CLA.

##### Member of the conference program committees

- Peggy Cellier: *CONCEPTS'24, ECMLPKDD'24, ECAI'24, EGC'24, IDA'24, FCA4AI'24*
- Romaric Gaudel: *NeurIPS'24 (A\*), ICML'24 (A\*), AAAI'24 (A\*), prix de thèse SSFAM'24, SSDBM'24, CAp'24, AIMLAI'24*
- Luis Galárraga: *The Web Conference (A\*), KR (A\*), ISWC (A), ESWC (A)*
- Paul Viallard: *NeurIPS'24 (A\*), ECMLPKDD'24 (A)*
- Sebastien Ferré: *IDA, CONCEPTS, FCA4AI*
- Alexandre Termier: *SDM'24, KDD'24, ECDM/PKDD'24, AIMLAI'24*
- Elisa Fromont: Area chair of *ECAI'24, KDD'24*, Senior PC of *ECDM/PKDD'24, IJCAI24* and *IDA'24*.
- Laurence Rozé: *ECAI'24*
- Christine Largouët: *ECAI'24, APIA'24*

### 11.1.3 Journal

#### Member of the editorial boards

- **Luis Galárraga** was co-guest editor (with Miguel Couceiro) of the Special Issue on **Fair and Explainable Decision Support Systems** published on February 2024 at the **EURO Journal of Decision Processes**. This special issue was a follow-up of a special session on the same topic organized at the EURO Conference 2022 by L. Galárraga and M. Couceiro.
- Alexandre Termier is a member of the editorial board of the Data Mining and Knowledge Discovery (DMKD) journal.

#### Reviewer - reviewing activities

- Sebastien Ferre: Semantic Web
- Romaric Gaudel: **Transactions on Machine Learning Research**
- Luis Galárraga: **InfoSys** (A+, 2024), **IPM** (Q1, 2024), **IDA** (Q3, 2024)
- Paul Viillard: **Journal of Machine Learning Research** (Q1, 2024)
- Alexandre Termier: **Data Mining and Knowledge Discovery (DMKD)**

### 11.1.4 Invited talks

- **Luis Galárraga** gave a course on **Explicabilité et “Machine Learning” Supervisé** (Explainability and Supervised Machine Learning) at the **MDD’24** (Masse de Données) Summer School organized by the **French Community on Data Management** on June 2024.
- **Paul Viillard** was invited for presenting his paper “Tighter Generalisation Bounds via Interpolation” to the **Journée MAS 2024**.
- Peggy Cellier and Sebastien Ferré were invited to give a talk on “Étude de l’utilisation des modèles de langages pour l’interrogation en langue naturelle des graphes de connaissances” at a series of seminar on **Geo-Historical Knowledge Graphs**, organized by GDR MAGIS, on 7 November 2024.
- Alexandre Termier was invited to give a talk on “Introduction to Generative AI” at the scientific committee of the PHASE department of Inrae, on the 23rd of May 2024. He also gave this talk at an internal seminar of Inrae Centre Val de Loire on the 27th of September 2024. Moreover, he presented the WAIT4 PEPR project at the annual meeting of the DigitAg project on the 23rd of April 2024.
- Elisa Fromont was Keynote (in French) on « Conférence sur l’I.A. », Auditorium de l’ESCC, Maen Roch, France. The 4/04/2024. She was panellist (in French) at the Techno-conférence on Generative AI organisée par le pôle de compétitivité I&R on the 10/10/2024. She was keynote speaker (in French) at Forum campus psy 2024 : Désir d’intelligence artificielle ? on "L’IA et le traitement de la parole", IFSI Rennes on the 12/10/2024. She was invited speaker (in French) at Les Territoriales de Bretagne on « IA et société : là où le bât blesse (et ce qu’on peut y faire) », Saint Briec on the 3/12/2024 and speaker and Panellist (in French) at the Séminaire DIGIT-BIO on "Data augmentation with style transfer", Lyon.
- **Christine Largouët** was invited by the Consulate General of France in Chicago to give a talk during the French American Innovation Day, Digital Ag & AI, on "Artificial Intelligence and Agriculture", 06/02/2024. She was keynote speaker (in French) at the seminar Ermerg’IN LiPh4SAS, INRAE on "Intelligence Artificielle pour le pilotage des élevages", 16/05/2024.

### 11.1.5 Leadership within the scientific community

- Elisa Fromont is in the steering committee of the International Conference **IDA** and of the French Conference on machine learning (CAp).

### 11.1.6 Scientific expertise

- Alexandre Termier reviewed a project for the Knowledge Foundation (Sweden), an application for an associate professor position at the University of Haifa (Israel), and an associate team proposal for Inria.
- Elisa Fromont As a member of the CSV ("Comité de Sélection et de Validation") of Images & Réseaux (pôle de compétitivité) she reviewed industrial projects all year long. She also reviewed project as a member of the scientific council of the MathNum dept of Inrae.
- Peggy Cellier was member of an associate professor recruitment committee in La Réunion (Université de La Réunion).
- Christine Largouët took part in the evaluation process of ANR-AAPG 2024 as a Scientific Expert for the scientific panel CE23 - Artificial intelligence and data science.

### 11.1.7 Research administration

- **Peggy Cellier** is in charge of the Phd students of the IRISA lab (commission personnel each month, etc). She was elected at the "Conseil de coposante" of the Computer Science departement of INSA Rennes until the end of 2024. She is also a member of "Conseil de l'école doctorale MATISSE".
- **Romarc Gaudel**: Romarc Gaudel is elected member of IRISA's board since dec. 2024.
- Elisa Fromont is the head of the D7 dpt of IRISA. She is elected at the scientific council of Université de Rennes and as such, she is also a member of the academic council (CAC) of the University and a member of the HDR Committee of the University. She is a member of the executive board of the CominLabs LABEX for the "Data, IA and Robotics" scientific axis.

## 11.2 Teaching - Supervision - Juries

### 11.2.1 Teaching

Apart from Luis Galárraga and Paul Viillard (research scientists), and Gaelle Tworkowski (administrative assistant), each permanent member of the project-team LACODAM is also faculty members and is actively involved in computer science teaching programs in ISTIC, IUT of Lannion, INSA, or Agrocampus-Ouest. Besides these usual teachings LACODAM is responsible of some teaching tracks and of some courses.

#### Teaching tracks responsibility

- **Véronique Masson** is the head of the L3 studies in Computer Science at University of Rennes
- **Alexandre Termier** is co-head of Master 2 SIF (Science Informatique - research master in Computer Science) at University of Rennes, with Matthieu Acher (INSA Rennes).
- **Sebastien Ferre** is the head of Master M1 Miage, and of the EIT international master track in Data Science (about 75 students).
- **Peggy Cellier** is the head of the last year at Computer Science Department at INSA (master 2 level, about 70 students).
- **Tassadit Bouadi** was head of continuation of studies at IUT of Lannion (computer science department), until July 2023. Since September 2023, she is co-head (with **Romarc Gaudel**) of the future Master M1 and M2 Artificial Intelligence at ISTIC, University of Rennes.
- **Christine Largouet** was head of the computer science educational unit at Institut Agro Rennes Angers (2 engineering schools) until septembre 2024. Since septembre 2023, she is co-head of the new master M1 and M2 E2C (Water, Energy and Climate, climate change mitigation and adaptation) at Institut Agro Rennes Angers.
- **Laurence Rozé** is the head of the L2 studies at INSA of Rennes (279 students).

### Courses responsibility

- **Alexandre Termier** is responsible for the following courses at ISTIC (Univ. Rennes): Object Programming (L2 info, elec, maths), Data Mining and Visualization (M2 SIF).
- **Elisa Fromont** is responsible of the "Deep Learning for Vision" (DLV) course (M2 SIF), the Machine Learning course (M2 IL) and teaches AI in M1 Info and L2 Info.
- **Luis Galárraga** (i) taught 6h within the course "Semantic Web" (by Sébastien Ferré M1 MIAGE ISTIC, Mar 2023); (ii) was responsible of 42h of teaching (22 TP + 20 TD) for the course "Java Programming" (INSA, Licence 1 INFO)
- **Peggy Cellier** is responsible of 5 courses at INSA Rennes: "Graphs and Algorithms" (Licence 3 INFO), "Databases" (Licence 3 Math), "Data Analysis and Data Mining" (Licence 3 INFO), "Advanced Database and Semantic Web" (Master 2) and "Ethique" (Master 2). She also teaches some other courses: "Use and functionalities of an operating system" (Licence 3). At master 2 SIF, she teaches in English 4,5 hours in the data mining course (DMV). In addition she gives a lecture of 2 hours also in master 2 SIF about "Qu'est-ce qu'une thèse, un doctorat, un-e doctorant-e?".
- **Sébastien Ferre** is responsible of 5 courses at ISTIC: "Basics of Data Analysis with Python" (M1 Miage EIT, in English), "Semantic Web Technologies" (M1 Miage, in English), "Data Mining" (M2 Miage, in English), "Compilers" (M1 info), "Technological Watch" (M1 Miage EIT).
- **Romarc Gaudel** is responsible for the following courses at ISTIC (Univ. Rennes): "discover AI" (L2), "Machine Learning" (M1 SIF) Data analysis and probabilistic modeling (M2 SIF), a course on recommender systems (M2 Miage & IET), a course on information retrieval and natural language processing (M2 Miage).
- **Tassadit Bouadi** is responsible for the following courses at IUT of Lannion (Univ. Rennes): SAé Creation of a database (BUT1 info) and Exploitation of a database (BUT1 info). And she is co-responsible of SQL and Programming course (BUT2 info). Since September 2023, she joined ISTIC, and is responsible of AI (M1 Info) course.
- **Christine Largouet** is responsible of the following courses at Institut Agro - Rennes Angers: Databases (L2 and L3), Programming in Python (L3), Scientific Programming (M1), Data Management and Machine Learning (M1), Artificial Intelligence (M2 E2C - Water Energy and Climate).
- **Laurence Rozé** is responsible of the following courses at INSA Rennes : probability (L3), mobile programming (L3,M1).
- **Elisa Fromont** is responsible of the following courses at ISTIC (Univ Rennes) : Introduction to Machine Learning (M1IA), option Machine Learning (M2IL), Deep Learning for vision (M2 SIF).

### Other responsibilities

- **Peggy Cellier** is in charge of the APC (Approche par compétences) development for the Computer Science Department. She is also part of the IDPE (Ingénieur diplômé par l'état) committee of INSA Rennes. She also represents INSA Rennes in the CMA (Compétence et Métier d'Avenir) IA TIARE.
- Alexandre Termier is an elected member of the Department Committee (conseil d'UFR) of the ISTIC departement of University of Rennes.
- Elisa Fromont is the scientific director of the CMA IA TIARE. She spends on average 1/2 days per weeks on this project: creation of new training programs (e.g. AI Master), scientific mediation, developpement of the continuous learning program, datalab, recruitments, ...)

## 11.2.2 Supervision

### PhD. Students

- (defended in 2024) **Olivier Gauriau**, (Inria, DigitAg, Acta Toulouse) 2021-2024; supervisors: Luis Galárraga, François Brun, Alexandre Termier and David Makowski; title: Numerical Rule Mining for the Prediction of the Dynamics of Crop Diseases, ED Matisse.
- (defended in 2024) **Elodie Germani**, 2021-2024; supervisors: Elisa Fromont and Camille Maumet (EMPENN); title: on representation learning for more robust FMRI data analysis, ED Matisse.
- **Gwladys Kelodjou**, 2022-2025; supervisors: Véronique Masson, Laurence Rozé, Alexandre Termier; title: Beyond the oracle: stabilizing the interpretability of machine learning algorithms, ED Matisse.
- **Charbel Kindji**, 2022-2025; supervisors: Elisa Fromont and Tanguy Urvoy (OrangeLabs); title: Architectures connexionnistes pour la génération de données tabulaires, ED Matisse.
- **Lucie Lepetit**, 2022-2025; supervisors: Peggy Cellier, Bruno Crémilleux and Alexandre Termier; title: Data mining methods for discovering behaviors related to animal well-being in precision farming data, ED Matisse.
- **Pierre Maurand**, 2022-2025; supervisors: Tassadit Bouadi, Peggy Cellier, Bruno Crémilleux (GREYC) and Alexandre Termier; title: Tell me your preferences and I will show you what you are interested in, ED Matisse.
- **Vanessa Fokou**, 2022-2025; supervisors: Florence Le Ber and Xavier Dolques (Univ. Strasbourg), Sebastien Ferre, Peggy Cellier; title: Comparison and cooperation of different Formal Concept Analysis approaches for relational data, Univ. Strasbourg.
- **Isseinie Sinouvassane**, 2023-2026; supervisors: Luis Galarraga, Alexandre Termier; title: How-Provenance Polynomials for Efficient and Greener Rule Mining, ED Matisse (financed by an ENS scholarship).
- **Yasmine Hachani**, 2023-2026; supervisors: Patrick Bouthémy (SAIRPICO), Elisa Fromont; title: Analyse par apprentissage profond de la dynamique du développement précoce des embryons bovins à partir de vidéomicroscopie, ED Matisse.
- **Paul Sevellec**, 2023-2026; supervisors: Matteo Sammarco (Stellantis), Elisa Fromont, Romaric Gaudel, Laurence Rozé ; title: Explications de séries temporelles multivariées par contrefactuels, ED Matisse.
- **Dimitri Lerévérénd**, 2023-2026; supervisors: Davide Frey (WIDE), Romaric Gaudel ; title: Privacy Preserving Decentralized Through Model Fragmentation, ED Matisse.
- **Niels Cobat**, 2024-2027; supervisors: Romaric Gaudel, Damien Hardy (PACAP) ; title: Analyse et optimisation des fichiers d'impression 3D à l'aide de méthodes d'apprentissage automatique, ED Matisse.
- **Ismail Bachchar**, 2024-2027; supervisors: Tassadit Bouadi, Thomas Guyet (AISTROSIGHT) ; title: Conception d'explications individuelles locales robustes, ED InfoMaths.
- **Sacha Germain**, 2024-2027; supervisors: Christine Largouët, Laurence Rozé, Tassadit Bouadi ; title: Detection et explications du comportement individuel et collectif des individus au sein d'un groupe pour estimer leur bien-être, ED Matisse.
- **Manuel Nkegoum**, 2024-2027; supervisors: Elisa Fromont, Sébastien Lefèvre (OBELIX) ; title: Object detection from few multispectral examples, ED Matisse.
- **Youwan Mahé**, 2024-2027; supervisors: Elisa Fromont, Elise Bannier (EMPENN) ; title: Détection et segmentation d'anomalies pour la caractérisation de la récupération post AVC, ED Matisse.
- **Nouha Karaouli**, 2024-2027; supervisors: Elisa Fromont, Denis Coquenet (SHADOC) ; title: Incremental Deep Learning for Embedded Systems, ED Matisse.

### Internships

- **Hind Atbir** (M2); supervisor: Paul Viillard, Farah Cherfaoui (Université de Saint-Etienne), Emilie Morvant (Université de Saint-Etienne), Guillaume Metzler (Université Lumière Lyon 2); title: PAC-Bayesian Fair Learning
- **Niels Cobat** (M2); supervisor: Romaric Gaudel, Damien Hardy (PACAP); title: Estimation précise du temps d'impression 3D par machine learning
- **Sacha Germain** (M2); supervisor: Christine Largouët (with Sébastien Picault, INRAE) ; title: Prediction of respiratory diseases occurrence in connected cattle farms with machine learning methods
- **Loane Portier** (L3); supervisor: Luis Galarraga Del Prado ; title: Multimodal Explainable Knowledge Graph Embeddings
- **Amina Lahlah** (L3); supervisor: Laurence Rozé, Véronique Masson ; title: Tester et compléter une nouvelle version du logiciel Leftist
- **Romain Eisenschmidt** (L3); supervisor: Elisa Fromont ; title: Dépollution de la science
- **Enzo Ouattara** (L3); supervisor: Laurence Rozé ; title: Analyse et correction des données de recensement des marins français : stratégies d'évaluation et construction d'un rapport interactif

### Engineer

- **Julianne Guerbette**, 2024 ; supervisor: Luis Galárraga; **Project:** Different tasks in eXplainable AI. Julianne worked mainly in the put in shape of the AMIE rule mining system including a refactoring of its codebase, the development of a server backend for the storage (database) layer, and the development and evaluation of lightweight estimations of the rule metrics for the sake of speed-up and scalability.
- **Frederic Lang**, 2024-2025; supervisors: Sebastien Ferré, Peggy Cellier; project: SmartFCA. Frédéric worked on the SmartFCA platform, and on Graph-FCA. He developed an OCaml version of part of the framework, and then used that to lift our Graph-FCA tool as a SmartFCA component that can be integrated into the platform.

### Apprentice

- **Loane Portier**, 2024-2026; supervisor: Elisa Fromont; Title: Post-hoc Explanations for Federated Learning Systems

### 11.2.3 PHD & HDR Juries

- **Peggy Cellier** was a member of the following PhD juries in 2024: Thomas Georges, 23/01 Univ. Montpellier (PhD, reviewer); Etienne Lehembre, 06/09 Univ. de Caen (PhD, examinatrice); Sarah Theroine, 20/12 Univ. Bourgogne (PhD, examinatrice).
- **Elisa Fromont** (14, 50% local): Rehan Juhboo, 10/01 Saint-Etienne (reviewer); Léonard Tschora, 17/01 Lyon (committee member, president); Lorenzo Perini 31/01 & 28/03, Leuven, Be (reviewer), Zoubida Ameer, 24/05, Rennes (president); Elodie Germani, Rennes, 16/09 (co-supervisor); Lincen Yang, Leiden, 20/09, NL (reviewer); Reda Marzouk, 17/10, Nantes; Malik Kazi Aoual, 20/11, Paris; Florent Imbert, 22/11, Rennes; Hasnaa Ouadoudi Belabzioui, 20/12, Rennes. **HDR** Rémi Emonet, 05/02, Saint-Etienne (president); Camille Maumet, 8/02, Rennes (president); Tristan Allard, 10/04, Rennes (president); Mathieu Lefort, 13/12, Lyon (committee member).
- **Sebastien Ferre** was a member of the following PhD juries in 2024: Sarah Ghidalia, 10/01 Univ. Bourgogne (rapporteur).
- **Alexandre Termier** was a member of the following PhD juries in 2024: Romain Brisse, 15/02 CentraleSupélec Rennes (examiner, president); Maxime Fucellaro, 25/11 Univ. Bordeaux (reviewer); Cedric Gernigon, 18/12 Univ. Rennes (examiner)
- **Romaric Gaudel** was a member of the following jury: Ludovic DELANDE, 25/09 University of Toulouse (PhD, reviewer);



#### 11.2.4 Doctoral advisory comitee (CSID)

- **Peggy Cellier** was a member of the mid-term evaluation committee of Oumaima El Khettari (Université de Nantes); Clémence Sebe (Université Paris Saclay); Randa Bendjeddou (Université de Lyon 2); Yacine Mokhtari (IMT Atlantique Brest) and Jules Berry (INSA Rennes, IRMAR).
- **Luis Galárraga** was a member of the mid-term evaluation committee of Ataollah Kamal (INSA Lyon) and Sacha Corbugy (Université de Namur), and Owen Le Godinec (University of Rennes)
- **Romaric Gaudel** was a member of the mid-term evaluation juries of Émile Binot.
- **Elisa Fromont** was a member of the mid-term evaluation juries of Ewan MOREL-CORLU (Rennes) 16/09/2024; Bruno Michelot (Lyon) 05/09/24; Loïc Eyango (Rennes) 14/06/2024; Irina Proskurina (Lyon) 2/05/2024; Ricky Walsh (Rennes) 3/04/2024.
- **Laurence Rozé** was a member of the mid-term evaluation juries of Nolwenn Pinczon du sel.
- **Christine Largouët** was a member of the mid-term evaluation juries of Maryem BEN SALEM (Université de Rennes) 21/06/2024.

### 11.3 Popularization

#### 11.3.1 Specific official responsibilities in science outreach structures

- **Tassadit Bouadi** is co-organizer of the project *L Codent, L Créent Rennes*, since 2018.
- **Elisa Fromont** is co-organizer of the project *J'peux pas j'ai informatique* for secondary school teacher, since 2022. She co-organzied a "JFMI" day for (100) secondary school female students in December.

#### 11.3.2 Productions (articles, videos, podcasts, serious games, ...)

- **Luis Galárraga** was co-author of the chapter **Neurosymbolic Methods for Rule Mining** of the book *The Handbook on Neurosymbolic AI and Knowledge Graphs* to be published by IOS Press series "Frontiers in Artificial Intelligence and Applications"

#### 11.3.3 Participation in Live events

- **Romaric Gaudel** did the presentation "Comment intégrer l'intelligence artificielle et ses enjeux à l'enseignement de l'informatique ?" during "Les rendez-vous de l'informatique" at Rennes to, 11/04 and 12/04.
- **Romaric Gaudel** did the presentation "IA & Société, Là où le bât blesse... (et ce qu'on peut y faire)" to Students at INSA Rennes, 15/10.
- **Elisa Fromont** on 21/02/2024, 15/03/2024, 5/04/2024 she gave a conference (in French) for secondary and high school students "Intelligence Artificielle de quoi parle-t-on ?". 2 conferences were done within the Chiche ! Inria program. On 19/03/24, she made a small videolecture for the 5 minutes of Lebesgue LABEX (in French) on "Les modèles de diffusion en apprentissage automatique". On 31/05/2024, 19/06/2024, 26/06/2024, she participated in training sessions in AI for Bachelor female Mathematicians involved in Math C pour L and (female-only) high school students. On 02/09/2024, she co-organized and gave an introductory conference at the AI summer school of the TIARe project on the "Societal Impacts of AI" in Rennes.
- **Peggy Cellier** gave a conference (in French) for Students at INSA "Intelligence Artificielle : quelques éléments de réflexion" in the context of "Concours d'éloquence 2024". She also gave a conference in french for the SAF (Service des Affaires Financières) team of Inria RBA about "Intelligence Artificielle".

### 11.3.4 Others science outreach relevant activities

- **Luis Galárraga** and **Christine Largouët** participated as experts guests at the episode “**Dans le secret du fonctionnement des IA**” of the scientific dissemination show *Science en Question* filmed on March 2024 and released on July 2024 by *L'Esprit Sorcier TV*

## 12 Scientific production

### 12.1 Major publications

- [1] V. Bellon Maurel, L. Brossard, F. Garcia, N. Mitton and A. Termier. *Agriculture and Digital Technology: Getting the most out of digital technology to contribute to the transition to sustainable agriculture and food systems*. Jan. 2022, pp. 1–185. DOI: [10.17180/wmkb-ty56-en](https://doi.org/10.17180/wmkb-ty56-en). URL: <https://hal.inrae.fr/hal-03604970>.
- [2] P. Cellier, M. Ducassé, S. Ferré, O. Ridoux and W. E. Wong. ‘Data Mining-Based Techniques for Software Fault Localization’. In: *Handbook of Software Fault Localization*. 1. Wiley, 20th Apr. 2023, Chapitre 7. DOI: [10.1002/9781119880929.ch7](https://doi.org/10.1002/9781119880929.ch7). URL: <https://hal.science/hal-04403506>.
- [3] S. Corbillé, E. Anquetil and E. Fromont. ‘Precise Segmentation for Children Handwriting Analysis by Combining Multiple Deep Models with Online Knowledge’. In: *ICDAR 2023 - 17th International Conference on Document Analysis and Recognition*. San José, United States, 21st Aug. 2023, pp. 1–18. URL: <https://hal.science/hal-04142592>.
- [4] L. Cornanguer, C. Largouët, L. Rozé and A. Termier. ‘TAG: Learning Timed Automata from Logs’. In: *AAAI 2022 - 36th AAAI Conference on Artificial Intelligence*. Virtual, Canada, 22nd Feb. 2022, pp. 1–9. URL: <https://hal.inria.fr/hal-03564455>.
- [5] M. Durand, C. Largouët, L. Bonneau de Beaufort, J.-Y. Dourmad and C. Gaillard. ‘Prediction of the daily nutrient requirements of gestating sows based on sensor data and machine-learning algorithms’. In: *Journal of Animal Science* 101 (2023), skad337. DOI: [10.1093/jas/skad337](https://doi.org/10.1093/jas/skad337). URL: <https://hal.inrae.fr/hal-04235874>.
- [6] K. Fauvel, E. Fromont, V. Masson, P. Faverdin and A. Termier. ‘XEM: An explainable-by-design ensemble method for multivariate time series classification’. In: *Data Mining and Knowledge Discovery* 36.3 (11th Feb. 2022), pp. 917–957. DOI: [10.1007/s10618-022-00823-6](https://doi.org/10.1007/s10618-022-00823-6). URL: <https://hal.inria.fr/hal-03599214>.
- [7] K. Fauvel, V. Masson, E. Fromont, P. Faverdin and A. Termier. ‘Towards Sustainable Dairy Management - A Machine Learning Enhanced Method for Estrus Detection’. In: *KDD 2019 - ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 25th SIGKDD Conference on Knowledge Discovery and Data Mining proceedings. Anchorage, United States, Aug. 2019, pp. 1–9. DOI: [10.1145/3292500.3330712](https://doi.org/10.1145/3292500.3330712). URL: <https://hal.archives-ouvertes.fr/hal-02190790>.
- [8] S. Felton, É. Fromont and E. Marchand. ‘Deep metric learning for visual servoing: when pose and image meet in latent space’. In: *ICRA 2023 - IEEE International Conference on Robotics and Automation*. London, United Kingdom: IEEE, 29th May 2023, pp. 741–747. DOI: [10.1109/ICRA48891.2023.10160963](https://doi.org/10.1109/ICRA48891.2023.10160963). URL: <https://inria.hal.science/hal-04003126>.
- [9] L. Galárraga, D. Hernández, A. Katim and K. Hose. ‘Visualizing How-Provenance Explanations for SPARQL Queries’. In: *WWW 2023 - ACM International World Wide Web Conference*. Austin, United States: ACM, 2023, pp. 212–216. DOI: [10.1145/3543873.3587350](https://doi.org/10.1145/3543873.3587350). URL: <https://inria.hal.science/hal-04386268>.
- [10] E. Galbrun, P. Cellier, N. Tatti, A. Termier and B. Crémilleux. ‘Mining Periodic Patterns with a MDL Criterion’. In: *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*. Dublin, Ireland, 2018. URL: <https://hal.archives-ouvertes.fr/hal-01951722>.

- [11] C.-S. Gauthier, R. Gaudel and E. Fromont. ‘UniRank: Unimodal Bandit Algorithm for Online Ranking’. In: ICML 2022 - 39th International Conference on Machine Learning. Baltimore, United States, 17th July 2022, pp. 1–31. URL: <https://hal.inria.fr/hal-03740981>.
- [12] C.-S. Gauthier, R. Gaudel, E. Fromont and B. A. Lompo. ‘Parametric Graph for Unimodal Ranking Bandit’. In: ICML 2021 - International Conference on Machine Learning. Vol. 139. Proceedings of the 38th International Conference on Machine Learning. Virtual, Canada, 2021, pp. 3630–3639. URL: <https://hal.archives-ouvertes.fr/hal-03256621>.
- [13] C. Gautrais, P. Cellier, T. Guyet, R. Quiniou and A. Termier. ‘Sky-signatures: detecting and characterizing recurrent behavior in sequential data’. In: *Data Mining and Knowledge Discovery* (29th Aug. 2023). DOI: [10.1007/s10618-023-00949-1](https://doi.org/10.1007/s10618-023-00949-1). URL: <https://hal.science/hal-04401641>.
- [14] E. Germani, E. Fromont and C. Maumet. ‘On the benefits of self-taught learning for brain decoding’. In: *GigaScience* 12 (3rd May 2023), pp. 1–17. DOI: [10.1093/gigascience/giad029](https://doi.org/10.1093/gigascience/giad029). URL: <https://inria.hal.science/hal-03769993>.
- [15] T. Guyet and R. Quiniou. ‘NegPSpan: efficient extraction of negative sequential patterns with embedding constraints’. In: *Data Mining and Knowledge Discovery* 34 (2020), pp. 563–609. DOI: [10.1007/s10618-019-00672-w](https://doi.org/10.1007/s10618-019-00672-w). URL: <https://hal.inria.fr/hal-03025572>.
- [16] V. Guyomard, F. Fessant, T. Guyet, T. Bouadi and A. Termier. ‘Generating robust counterfactual explanations’. In: ECML/PKDD - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Turin (Italie), Italy, 2023, pp. 1–16. URL: <https://hal.science/hal-04255500>.
- [17] V. Guyomard, F. Fessant, T. Guyet, T. Bouadi and A. Termier. ‘Interactive Visualization of Counterfactual Explanations for Tabular Data’. In: ECML/PKDD 2023 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Vol. 14175. Lecture Notes in Computer Science. Turin, Italy: Springer Nature Switzerland, 17th Sept. 2023, pp. 330–334. DOI: [10.1007/978-3-031-43430-3\\_25](https://doi.org/10.1007/978-3-031-43430-3_25). URL: <https://hal.science/hal-04255496>.
- [18] J. Lehmann, P. Gattogi, D. Bhandiwad, S. Ferré and S. Vahdati. ‘Language Models as Controlled Natural Language Semantic Parsers for Knowledge Graph Question Answering’. In: *Frontiers in Artificial Intelligence and Applications*. ECAI 2023 - 26th European Conference on Artificial Intelligence. Vol. 372. Frontiers in Artificial Intelligence and Applications. Krakow (Cracovie), Poland: IOS Press, 28th Sept. 2023, pp. 1348–1356. DOI: [10.3233/FAIA230411](https://doi.org/10.3233/FAIA230411). URL: <https://inria.hal.science/hal-04269089>.
- [19] G. G. Méndez, L. Galárraga, K. Chiluiza and P. Mendoza. ‘Impressions and Strategies of Academic Advisors When Using a Grade Prediction Tool During Term Planning’. In: CHI 2023 - Conference on Human Factors in Computing Systems. Hamburg, Germany: ACM, 2023, pp. 1–18. DOI: [10.1145/3544548.3581575](https://doi.org/10.1145/3544548.3581575). URL: <https://inria.hal.science/hal-04132566>.

## 12.2 Publications of the year

### International journals

- [20] T. Allard, H. Asghar, G. Avoine, C. Bobineau, P. Cauchois, E. Fromont, A. Monreale, F. Naretto, R. Pellungrini, F. Pratesi, M.-C. Rousset and A. Voyez. ‘Analyzing and explaining privacy risks on time series data: ongoing work and challenges’. In: *SIGKDD explorations : newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining* 26.1 (2024), pp. 1–10. DOI: [10.1145/3682112.3682118](https://doi.org/10.1145/3682112.3682118). URL: <https://hal.science/hal-04628000>. In press (cit. on p. 17).
- [21] J. Delaunay, L. Galárraga and C. Largouët. ‘Expliquer une boîte noire sans boîte noire’. In: *Revue TAL : traitement automatique des langues* 64.3/2023 (2024), pp. 93–117. URL: <https://hal.science/hal-04744191> (cit. on p. 13).
- [22] O. Gauriau, L. Galárraga, F. Brun, A. Termier, L. Davadan and F. Joudelat. ‘Comparing machine-learning models of different levels of complexity for crop protection: A look into the complexity-accuracy tradeoff’. In: *Smart Agricultural Technology* 7 (Mar. 2024), p. 100380. DOI: [10.1016/j.atech.2023.100380](https://doi.org/10.1016/j.atech.2023.100380). URL: <https://hal.science/hal-04382202> (cit. on p. 16).

- [23] S. Moia, H.-T. Wang, A. S. Heinsfeld, D. Jarecka, Y. F. Yang, S. Heunis, M. Svanera, B. D. Leener, A. Gondová, S. Kim et al. 'Proceedings of the OHBM Brainhack 2022'. In: *Aperture Neuro* (18th Mar. 2024). DOI: [10.52294/001c.92760](https://doi.org/10.52294/001c.92760). URL: <https://inria.hal.science/hal-04478342>.

#### International peer-reviewed conferences

- [24] A. N. Amalou, E. Fromont and I. Puaut. 'Fast and Accurate Context-Aware Basic Block Timing Prediction using Transformers'. In: *Proceedings of the ACM SIGPLAN 2024 International Conference on Compiler Construction*. CC 2024 - ACM SIGPLAN 33rd International Conference on Compiler Construction. Edimbourg, United Kingdom: ACM; ACM, 2nd Mar. 2024, pp. 227–237. DOI: [10.1145/3640537.3641572](https://doi.org/10.1145/3640537.3641572). URL: <https://hal.science/hal-04406073> (cit. on p. 17).
- [25] Z. Asma, D. Hernández, L. Galárraga, G. Flouris, I. Fundulaki and K. Hose. 'NPCS: Native Provenance Computation for SPARQL'. In: WWW 2024 - ACM Web Conference. Singapore, Singapore: ACM, 13th May 2024, pp. 2085–2093. DOI: [10.1145/3589334.3645557](https://doi.org/10.1145/3589334.3645557). URL: <https://inria.hal.science/hal-04881765> (cit. on p. 12).
- [26] S. Ferré. 'Tackling the Abstraction and Reasoning Corpus (ARC) with Object-centric Models and the MDL Principle'. In: *LNAI, Springer*. IDA 2024 - Symposium on Intelligent Data Analysis. Stockholm, Sweden, 2024, pp. 1–12. URL: <https://inria.hal.science/hal-04602995> (cit. on p. 11).
- [27] E. Germani, E. Fromont and C. Maumet. 'Uncovering communities of pipelines in the task-fMRI analytical space'. In: IEEE International Conference on Image Processing. Abu Dhabi, United Arab Emirates, 27th Oct. 2024. DOI: [10.1109/ICIP51287.2024.10647701](https://doi.org/10.1109/ICIP51287.2024.10647701). URL: <https://hal.science/hal-04331232> (cit. on p. 15).
- [28] Y. Hachani, P. Bouthemy, E. Fromont, S. Ruffini, L. Laffont and A. de Paula Reis. 'Early prediction of the transferability of bovine embryos from videomicroscopy'. In: ICIIP 2024 - IEEE International Conference on Image Processing. Abu DHABI, United Arab Emirates: IEEE, 2024, pp. 1–6. URL: <https://hal.science/hal-04880222> (cit. on p. 16).
- [29] G. Kelodjou, L. Rozé, V. Masson, L. Galárraga, R. Gaudel, M. Tchuente and A. Termier. 'Shaping Up SHAP: Enhancing Stability through Layer-Wise Neighbor Selection'. In: AAI 2024 - 38th Annual AAI Conference on Artificial Intelligence. Vancouver, Canada, 2024, pp. 1–10. URL: <https://inria.hal.science/hal-04413022> (cit. on p. 14).
- [30] T. Lebrun, L. Béziaud, T. Allard, A. Boutet, S. Gambs and M. Maouche. 'Synthetic Data: Generate Avatar Data on Demand'. In: The International Web Information Systems Engineering conference (WISE). Vol. 15440. Lecture Notes in Computer Science. Doha-Qatar, France: Springer Nature Singapore, 27th Nov. 2025, pp. 193–203. DOI: [10.1007/978-981-96-0576-7\\_15](https://doi.org/10.1007/978-981-96-0576-7_15). URL: <https://hal.science/hal-04715055> (cit. on p. 14).
- [31] A. É. Padonou, P. Cellier and S. Ferré. 'Investigating the use of language models for natural language querying over knowledge graphs'. In: *RNTI*. EGC 2024 - 24ème conférence francophone sur l'Extraction et la Gestion des Connaissances. Vol. RNTI-E-40. Dijon, France, 2024, pp. 1–8. URL: <https://inria.hal.science/hal-04602983> (cit. on p. 13).
- [32] J. Patracone, P. Viallard, E. Morvant, G. Gasso, A. Habrard and S. Canu. 'A Theoretically Grounded Extension of Universal Attacks from the Attacker's Viewpoint'. In: ECML PKDD 2024 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Vilnius, Lithuania, 2024, pp. 1–27. DOI: [10.1007/978-3-031-70359-1\\_17](https://doi.org/10.1007/978-3-031-70359-1_17). URL: <https://hal.science/hal-03615461> (cit. on p. 14).
- [33] M. R. R. Qureshi, N. Es-Sebbani, L. Galárraga, Y. Graham, M. Couceiro and Z. Bouraoui. 'REFINE-LM: Mitigating Language Model Stereotypes via Reinforcement Learning'. In: *Frontiers in Artificial Intelligence and Applications*. ECAI 2024 - 27th European Conference on Artificial Intelligence. Vol. 392. Santiago de Compostela, Spain, 25th Oct. 2024, pp. 4027–4034. DOI: [10.3233/FAIA240970](https://doi.org/10.3233/FAIA240970). URL: <https://hal.science/hal-04872144> (cit. on p. 17).

- [34] P. Viillard, R. Emonet, A. Habrard, E. Morvant and V. Zantedeschi. ‘Leveraging PAC-Bayes Theory and Gibbs Distributions for Generalization Bounds with Complexity Measures’. In: *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. AISTATS 2024 - 27th International Conference on Artificial Intelligence and Statistics. Valencia, Spain, 2024. URL: <https://hal.science/hal-04473777> (cit. on p. 15).

#### National peer-reviewed Conferences

- [35] V. Fokou, P. Cellier, X. Dolques, S. Ferré and F. Le Ber. ‘Comparing Relational Concept Analysis and Graph-FCA on their Common Ground’. In: *Concepts 2024*, Cadiz, Spain, septembre 2024. Cadiz, Spain, Sept. 2024. URL: <https://hal.science/hal-04622856> (cit. on p. 12).
- [36] Y. Hachani, P. Bouthemy, S. Ruffini, L. Laffont, E. Fromont and A. d. P. Reis. ‘Prédiction précoce de la transférabilité d’embryons bovins par vidéomicroscopie’. In: *RFIAP 2024*. RFIAP 2024 - Congrès Reconnaissance des Formes, Image, Apprentissage et Perception. Lille, France, 2024. URL: <https://hal.science/hal-04614044> (cit. on p. 16).

#### Conferences without proceedings

- [37] H. Atbir, F. Cherfaoui, G. Metzler, E. Morvant and P. Viillard. ‘Une borne PAC-Bayésienne sur une mesure de risque pour l’apprentissage équitable’. In: *CAP 2024 - Conférence sur l’Apprentissage Automatique*. Lille, France, 2024. URL: <https://hal.science/hal-04646194> (cit. on p. 13).
- [38] G. C. N. Kindji, L. M. Rojas-Barahona, E. Fromont and T. Urvoy. ‘Cross-table Synthetic Tabular Data Detection’. In: *COLING 2025 Workshop on Detecting AI Generated Content*. Abu dahbi, United Arab Emirates, 19th Jan. 2025. URL: <https://hal.science/hal-04840865>.

#### Scientific book chapters

- [39] A. Ławrynowicz, L. Galárraga, M. Alam, B. Jaulmes, V. Zeman and T. Kliegr. ‘Neurosymbolic Methods for Rule Mining’. In: *Handbook on Neurosymbolic Artificial Intelligence*. 2024. URL: <https://hal.science/hal-04792154>. In press (cit. on p. 11).

#### Edition (books, proceedings, special issue of a journal)

- [40] *Conceptual Knowledge Structures*. CONCEPTS 2024 - 1st International Joint Conference on Conceptual Knowledge Structures. Vol. 14914. Lecture Notes in Computer Science. Springer Nature Switzerland, 2024, pp. 1–348. DOI: [10.1007/978-3-031-67868-4](https://doi.org/10.1007/978-3-031-67868-4). URL: <https://inria.hal.science/hal-04885540> (cit. on p. 12).
- [41] *SSDBM '24: Proceedings of the 36th International Conference on Scientific and Statistical Database Management*. SSDBM 2024 - 36th International Conference on Scientific and Statistical Database Management. ACM, 23rd Aug. 2024, pp. 1–165. DOI: [10.1145/3676288](https://doi.org/10.1145/3676288). URL: <https://hal.science/hal-04813517> (cit. on p. 12).

#### Doctoral dissertations and habilitation theses

- [42] O. Gauriau. ‘Numerical rule mining for the prediction of plant disease dynamics.’ Université de Rennes, 13th Nov. 2024. URL: <https://theses.hal.science/tel-04823559>.

#### Reports & preprints

- [43] B. Dupuis, P. Viillard, G. Deligiannidis and U. Şimşekli. *Uniform Generalization Bounds on Data-Dependent Hypothesis Sets via PAC-Bayesian Theory on Random Sets*. 26th Apr. 2024. URL: <https://hal.science/hal-04562426> (cit. on p. 13).
- [44] E. Germani, N. Baghwat, M. Dugré, R. Gau, A. Montillo, K. Nguyen, A. Sokolowski, M. Sharp, J.-B. Poline and T. Glatard. *Predicting Parkinson’s disease trajectory using clinical and functional MRI features: a reproduction and replication study*. 13th Sept. 2024. URL: <https://inserm.hal.science/inserm-04465765> (cit. on p. 15).

- [45] E. Germani, E. Fromont, P. Maurel and C. Maumet. *The HCP multi-pipeline dataset: an opportunity to investigate analytical variability in fMRI data analysis*. 2024. URL: <https://inserm.hal.science/inserm-04356768> (cit. on p. 15).
- [46] E. Germani, C. Maumet and E. Fromont. *Mitigating analytical variability in fMRI results with style transfer*. 2024. URL: <https://inserm.hal.science/inserm-04531405> (cit. on p. 16).
- [47] M. Haddouche, P. Viallard, U. Şimşekli and B. Guedj. *A PAC-Bayesian Link Between Generalisation and Flat Minima*. 13th Feb. 2024. URL: <https://hal.science/hal-04455639> (cit. on p. 14).
- [48] G. C. N. Kindji, L. M. Rojas-Barahona, E. Fromont and T. Urvoy. *Under the Hood of Tabular Data Generation Models: Benchmarks with Extensive Tuning*. 2024. URL: <https://hal.science/hal-04612244>.
- [49] P. Viallard, M. Haddouche, U. Şimşekli and B. Guedj. *Tighter Generalisation Bounds via Interpolation*. 2024. DOI: [10.48550/arXiv.2402.05101](https://doi.org/10.48550/arXiv.2402.05101). URL: <https://hal.science/hal-04456925> (cit. on p. 14).

### Other scientific publications

- [50] Y. Hachani, P. Bouthemy, S. Ruffini, L. Laffont, E. Fromont and A. de Paula Reis. ‘Early prediction of the transferability of bovine embryos from videomicroscopy’. In: *IABM 2024 - Colloque Français d’Intelligence Artificielle en Imagerie Biomédicale*. Grenoble, France, 2024, pp. 1–1. URL: <https://hal.science/hal-04880124> (cit. on p. 16).

### Scientific popularization

- [51] F. Dégeilh, E. Germani, N. Jégou, C. Joud, J. Lefort-Besnard and C. O. Muller. *À la découverte des métiers de la recherche en imagerie cérébrale*. Rennes, France, 14th Mar. 2024. URL: <https://inria.hal.science/hal-04574796>.

## 12.3 Cited publications

- [52] K. P. Nguyen, V. Raval, A. Treacher, C. Mellema, F. F. Yu, M. C. Pinho, R. M. Subramaniam, R. B. Dewey and A. A. Montillo. ‘Predicting Parkinson’s disease trajectory using clinical and neuroimaging baseline measures’. In: *Parkinsonism and Related Disorders* 85 (2021), pp. 44–51. DOI: <https://doi.org/10.1016/j.parkreldis.2021.02.026>. URL: <https://www.sciencedirect.com/science/article/pii/S1353802021000754> (cit. on p. 16).