

2025 Activity Report

RESEARCH CENTRE: Inria Centre at Rennes University

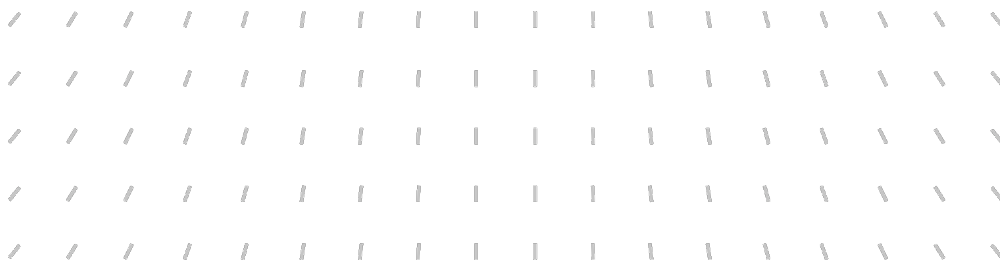
IN PARTNERSHIP WITH: Institut national des sciences appliquées de Rennes,
Institut national supérieur des sciences agronomiques, agroalimentaires, horticoles
et du paysage, Université de Rennes

Project-Team

LACODAM

Large scale Collaborative Data Mining

In collaboration with Institut de recherche en informatique et systèmes aléatoires
(IRISA)



Project-Team LACODAM

Creation of the Project-Team: 2017 November 01

Each year, Inria research teams publish an Activity Report presenting their work and results over the reporting period. These reports follow a common structure, with some optional sections depending on the specific team. They typically begin by outlining the overall objectives and research programme, including the main research themes, goals, and methodological approaches. They also describe the application domains targeted by the team, highlighting the scientific or societal contexts in which their work is situated. The reports then present the highlights of the year, covering major scientific achievements, software developments, or teaching contributions. When relevant, they include sections on software, platforms, and open data, detailing the tools developed and how they are shared. A substantial part is dedicated to new results, where scientific contributions are described in detail, often with subsections specifying participants and associated keywords. Finally, the Activity Report addresses funding, contracts, partnerships, and collaborations at various levels, from industrial agreements to international cooperations. It also covers dissemination and teaching activities, such as participation in scientific events, outreach, and supervision. The document concludes with a presentation of scientific production, including major publications and those produced during the year.

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
- A5.1. – Human-Computer Interaction
- A5.2. – Data visualization
- A5.3. – Image processing and analysis
 - A5.3.2. – Sparse modeling and image representation
- A9.1. – Knowledge
 - A9.2.1. – Supervised learning
 - A9.2.2. – Unsupervised learning
 - A9.2.4. – Optimization and learning
 - A9.2.5. – Bayesian methods
 - A9.2.6. – Neural networks
 - A9.2.8. – Deep learning
- A9.4. – Natural language processing
- A9.6. – Decision support
- A9.7. – AI algorithmics
- A9.8. – Reasoning
- A9.10. – Hybrid approaches for AI
- A9.11. – Generative AI
 - A9.12.1. – Object recognition
- A9.13. – Agentic AI
- A9.15. – Symbolic 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

Contents

Project-Team LACODAM	1
1 Team members, visitors, external collaborators	6
2 Overall objectives	7
3 Research program	7
3.1 Introduction	7
3.2 Symbolic methods	8
3.3 Interpretable ML	8
3.4 Real world AI	9
4 Application domains	9
4.1 Industry	9
4.2 Agriculture and Environment	10
4.3 Cognitive Sciences	11
4.4 Semantic Data Management	11
5 Social and environmental responsibility	11
5.1 Footprint of research activities	11
5.2 Impact of research results	12
6 Highlights of the year	12
7 Latest software developments, platforms, open data	12
7.1 Latest software developments	12
7.1.1 HIPAR	12
7.1.2 Dexteris	12
7.1.3 skm	13
7.2 New platforms	13
7.2.1 SmartFCA platform	13
7.3 Open data	13
8 New results	14
8.1 Symbolic Methods	14
8.1.1 Graph-FCA	14
8.1.2 Semantic Web	15
8.2 Interpretable Machine Learning	15
8.3 Real-World ML	16
8.3.1 Machine Learning on Sequences	16
8.3.2 Privacy and Machine Learning	17
8.3.3 Other Applications	18
9 Bilateral contracts and grants with industry	18
9.1 Bilateral contracts with industry	18
10 Partnerships and cooperations	19
10.1 International research visitors	19
10.1.1 Visits of international scientists	19
10.2 National initiatives	19
10.2.1 ANR	21

11 Dissemination	23
11.1 Promoting scientific activities	23
11.1.1 Scientific events: organisation	23
11.1.2 Scientific events: selection	23
11.1.3 Journal	23
11.1.4 Invited talks	24
11.1.5 Leadership within the scientific community	24
11.1.6 Scientific expertise	24
11.1.7 Research administration	24
11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach	24
11.2.1 Supervision	26
11.2.2 Juries	27
11.2.3 Doctoral advisory committee (CSID)	27
11.2.4 Educational and pedagogical outreach	27
11.3 Popularization	28
11.3.1 Participation in Live events	28
11.3.2 Others science outreach relevant activities	28
12 Scientific production	28
12.1 Major publications	28
12.2 Publications of the year	30
12.3 Cited publications	32

1 Team members, visitors, external collaborators

Research Scientists

- Luis Galarraga Del Prado [INRIA, Researcher]
- Gonzalo Mendez Cobena [INRIA, Starting Research Position, from Aug 2025 until Sep 2025]
- Gonzalo Mendez Cobena [INRIA, Starting Research Position, from May 2025 until May 2025]
- Gonzalo Mendez Cobena [INRIA, Starting Research Position, until Jan 2025]
- Paul Viallard [INRIA, ISFP, until Feb 2025]

Faculty Members

- Alexandre Termier [Team leader, UNIV RENNES, Professor, HDR]
- Tassadit Bouadi [UNIV RENNES, Associate Professor]
- Peggy Cellier [INSA RENNES, Associate Professor, HDR]
- Sebastien Ferre [UNIV RENNES, Professor, until Aug 2025, HDR]
- Elisa Fromont [UNIV RENNES, Professor, until Feb 2025, HDR]
- Romaric Gaudel [UNIV RENNES, Associate Professor, until Feb 2025, HDR]
- Christine Largouet [L'INSTITUT AGRO, Professor, HDR]
- Véronique Masson [UNIV RENNES , Associate Professor]
- Laurence Rozé [INSA RENNES, Associate Professor]

Post-Doctoral Fellow

- Aurélien Lamercherie [UNIV RENNES]

PhD Students

- Ismail Bachchar [Orange Labs, CIFRE, with AISTROSIGHT team]
- Sacha Germain [INRIA]
- Julianne Guerbette [UNIV RENNES , from Feb 2025, with MALT team]
- Gwladys Kelodjou [UNIV RENNES]
- Lucie Lepetit [INRIA, until Jul 2025]
- Pierre Maurand [INSA RENNES, until Feb 2025]
- Paul Sevellec [Stellantis, University of Rennes, CIFRE, with MALT team]
- Isseinie Sinouvassane [UNIV RENNES]

Technical Staff

- Louis Bonneau De Beaufort [L'INSTITUT AGRO, Engineer]
- Pierre Cottais [INRIA, Engineer, from Mar 2025]
- Marine Hamon [INRIA, Engineer, from Mar 2025]
- Frederic Lang [UNIV RENNES, Engineer]

Interns and Apprentices

- Lydia Achour [INRIA, Intern, from May 2025 until Aug 2025]
- Wissam Aissaoui [UNIV RENNES, Intern, from Dec 2025]
- Baptiste Amice [UNIV RENNES, Intern, from Feb 2025 until Jul 2025]
- Maxime Desbans [UNIV RENNES, Intern, from Dec 2025]
- Isidore Gomendy [INRIA, Intern, from Jun 2025 until Jul 2025]

Administrative Assistant

- Gaelle Tworkowski [INRIA]

External Collaborator

- Gonzalo Mendez Cobena [Universitat Politècnica de València, from Feb 2025, Three different stays: Feb-Apr, Jun-Jul, Oct-Dec]

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) [44]. 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 [43]. 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) [46, 45], 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 Kinji (now at MALT), 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.

- **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 (now at MALT), we explore the potential of Machine Learning models dedicated to sequences to handle this tasks.
- **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 is a key component of these improvements. This is the general topic of the WAIT4 project (see Section 10.2), 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, tackled by a collaboration between our engineers Marine Hamon and Pierre Cottais, 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 Lénaig Cornanguer, which defended her PhD in the Lacodam team in 2023.
- **Deep learning-based analysis of the early development of bovine embryos from videomicroscopy.** The PhD of Yasmine Hachani (now at MALT, 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 Del Prado and Gonzalo Mendez Cobena), 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 operations¹. 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).

¹Processes where there is some sort of data difference

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 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

An important event this year has been the creation of the MALT team in March 2025, with three former members of Lacodam: Elisa Fromont (MALT team leader), Romaric Gaudel and Paul Viillard. Lacodam and MALT are now two teams with different organisations and different research interests (with some overlaps), but we still exchange on a daily basis and organise some joint events (joint seminars or convivial events). Lacodam members have submitted a research project for a new team, HYWOKX, which is undergoing the Inria review process.

Christine Largouët has been promoted to Full Professor at Institut Agro Rennes-Anger.

7 Latest software developments, platforms, open data

7.1 Latest software developments

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, i- terations, 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](#)

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

7.2 New platforms

7.2.1 SmartFCA platform

Name: The SmartFCA platform

Keywords: Formal concept analysis, Graph-FCA

Functional Description

The SmartFCA platform is a micro-services based platform. Several services are working together in order to achieve complex computations. Services can be used through a graphical user interface (Web application), or you can also directly send request to separated RESTFUL APIs.

Contact: Frédéric Lang

Participants: Peggy Cellier, Sébastien Ferré, Frédéric Lang.

URL: [Link to the SmartFCA platform](#)

7.3 Open data

HotPig: A Behavioural Dataset of Pigs under Heat Stress

Contributors: Louis Bonneau de Beaufort, Xavier Caroline, David Renaudeau, Christine Largouët, Florence Gondret

Description: The widespread use of videos in modern indoor livestock facilities coupled with the availability of efficient and low-cost computer vision algorithms provides strong incentives for continuously monitoring farm animal behaviour. Deciphering how pigs behave when experiencing prolonged heat stress (HS) is particularly important for animal welfare, as it helps us to better understand how animals use various thermoregulation and heat dissipation mechanisms. This dataset includes the monitoring of continuous behavioural traits for 24 growing pigs first housed at thermoneutrality and then exposed to HS. The data can be used to illustrate the frequencies of specific behavioural traits (time budget)

and their deviations due to heat stress, either on average or in animal-centred view (recurrence of patterns, etc.). Outputs can be used to perform behavioural patterns mining, behaviour clustering and modelling. An important effort was made to ensure consistency of the behavioural dataset, with comparison with readings of automatic feeders to decipher feeding visits vs. non-feeding visits. Further video processing algorithms may benefit from the training (labelled images) dataset, but also from the multiple annotation approach (postures and events). This dataset can be used to train any machine learning methods for behaviour prediction from videos in conventional growing pigs.

Data were collected on 24 pigs that were video-monitored day and night under two contrasted conditions: thermoneutral (TN, 22°C) and HS (32°C). All pigs were housed individually and had free access to an automatic electronic feeder delivering pellets four times a day, and to water. Environmental conditions (temperature, humidity) in the room were recorded by sensors. After acquisition, videos were processed using YOLOv11, a real-time object detection algorithm that uses a convolutional neural network (CNN), to extract the following behavioural traits: drinking, willingness to eat, lying down, standing up, moving around, curiosity towards the littermate housed in the neighbouring pen, and contact between the two animals (cuddling). A minute frequency sampling rate was applied (each minute correspond to 150 frames processed) for a continuous period of 16 days, spanning the two different thermal conditions (9 days on TN, 6 days on HS, 1 day back to TN). The algorithm was first trained thanks to manual video analysis labelling at the individual scale. Consistency with the automatic electronic feeder's data (also provided) was thoroughly checked. The dataset allows quantitative criterion to be analysed to decipher inter-individual differences in animal behaviour and their dynamic adaptation to heat stress.

Dataset PID (DOI,...): [10.5281/zenodo.15608129](https://doi.org/10.5281/zenodo.15608129)

Project link: <https://www.pepr-agroekonum.fr/les-projets-finances/agroequipements/projets-cibles/wait4>

Publications: [21]

Contact: Louis Bonneau de Beaufort

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). Some results may fall within several axes. In such cases we organize the result in its primary axis.

8.1 Symbolic Methods

Participants: Tassadit Bouadi, Peggy Cellier, Sébastien Ferré, Luis Galárraga, Alexandre Termier, Aurélien Lamercerie.

8.1.1 Graph-FCA

Conceptual Knowledge Structures. This book [40] 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.

Theoretical comparison of Relational Concept Analysis (RCA) and Graph-FCA (GCA) Relational Concept Analysis (RCA) and Graph-FCA (GCA) are two extensions of Formal Concept Analysis (FCA) introduced in order to allow concept analysis on multi-relational data [23]. The two methods have different properties and parameters, but when restricting to binary relationships, existential quantifier and unary concepts, their outputs look similar. On this basis, a theoretical comparison of the two methods is conducted, showing that each RCA concept corresponds to a GCA concept. Furthermore, to allow the comparison of concept intensions, a transformation of RCA results into relational patterns is performed. These results give a sound basis to help interpreting RCA results and to combine the two approaches for data exploration.

8.1.2 Semantic Web

Web-SPARQL: Hybrid Querying over Knowledge Graphs, Web, and Microdata. This paper [32] addresses the problem of querying semantic data from heterogeneous web sources. On one hand, centralized knowledge graphs, such as RDF stores, can be accessed with flexibility and efficiency using SPARQL queries. On the other hand, distributed knowledge graphs, such as microdata, are not directly queryable, and are rather exploited by search engines. RDF stores and microdata provide complementary information: RDF stores typically offer higher-quality data, while microdata delivers fresher content. The research problem considered in this paper is the hybrid querying of a centralized RDF store and distributed microdata on the web. To this aim, we introduce Web-SPARQL, an extension of SPARQL with property functions that link the centralized entities to the distributed entities on the web.

SparqLLM : Retrieval-Augmented SPARQL Query Processing. SPARQL is essential for querying Knowledge Graphs (KGs), but much information exists in external sources rather than within KGs. To address this, we propose SparqLLM, a retrieval-augmented query processing approach that leverages user-defined functions (UDFs) and named graphs to augment SPARQL queries with diverse external sources, including search engines, large language models (LLMs), and vector search. By doing so, SparqLLM significantly enhances SPARQL's capabilities, enabling a single query to access multiple heterogeneous sources while ensuring query provenance and explainability. This demonstration highlights the potential of SparqLLM to enrich query results with comprehensive, up-to-date information and showcases its application in a real-world use case [38].

Neurosymbolic Methods for Rule Mining. In this book chapter [39], 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.

8.2 Interpretable Machine Learning

Participants: Tassadit Bouadi, Julien Delaunay, Luis Galárraga, Romaric Gaudel, Gwladys Kelodjou, Christine Largouët, Véronique Masson, Laurence Rozé, Alexandre Termier, Paul Sevellec.

Generating Efficiently Realistic Counterfactual Explanations. This work presents VCNet (Variational CounterNet) [25], a method designed to generate realistic counterfactual explanations for tabular data, along with its extension ImmutableVCNet, which accounts for immutable features. VCNet aims to produce counterfactuals that are representative of their target classes while addressing key limitations of existing post-hoc and optimization-based approaches, notably high computational costs and suboptimal validity rates. Although several state-of-the-art methods mitigate these issues, they often generate counterfactuals that lack realism. VCNet addresses this shortcoming by incorporating explicit realism constraints into the generation process. The proposed approach relies on a conditional variational autoencoder (cVAE) that jointly models

the class-conditional data distributions, ensuring that generated counterfactuals both lie within the data manifold and are consistent with the target class distribution. ImmutableVCNet further extends this framework by enabling the handling of immutable features. Extensive ablation studies were conducted to analyze the impact of architectural design choices within VCNet. In addition, empirical evaluations demonstrate the effectiveness of the proposed methods in generating realistic counterfactuals. VCNet is evaluated against ImmutableVCNet, and ImmutableVCNet is compared with several state-of-the-art counterfactual generation methods.

Stratum by Stratum: Building Stable SHAP Explanations through Layered Approximations. SHAP is a popular post-hoc explainability method that assigns feature attribution scores based on the Shapley value from cooperative game theory. Since computing the exact SHAP values is intractable for large feature sets, several sampling-based approximation methods, such as KernelSHAP, have been proposed in the literature. These methods overcome intractability but suffer from stability issues. To address this instability, we propose the StratoSHAP family of stable SHAP approximations based on feature coalitions organized into strata. We provide formal analytical formulations for these approximations and demonstrate that they respect important properties of attribution values, such as stability, linearity, efficiency, symmetry, and fair treatment. A series of comparative experiments reveal that our stratum-based approach offers an interesting trade-off between computational complexity and approximation quality while remaining fully stable [42].

Impact of Explanation Technique and Representation on Users' Comprehension and Confidence in Explainable AI. Local explainability, an important sub-field of eXplainable AI, focuses on describing the decisions of AI models for individual use cases by providing the underlying relationships between a model's inputs and outputs. While the machine learning community has made substantial progress in improving explanation accuracy and completeness, these explanations are rarely evaluated by the final users. In this paper [22], we evaluate the impact of various explanation and representation techniques on users' comprehension and confidence. Through a user study on two different domains, we assessed three commonly used local explanation techniques: feature-attribution, rule-based, and counterfactual-and explored how their visual representation—graphical or text-based—influences users' comprehension and trust. Our results show that the choice of explanation technique primarily affects user comprehension, whereas the graphical representation impacts user confidence.

8.3 Real-World ML

Participants: Élisabeth Fromont, Luis Galárraga, Antonin Voyez, Laurence Rozé, Paul Sevellec, Gonzalo Méndez, Gaspard Kindji, Elodie Germani.

8.3.1 Machine Learning on Sequences

Plausible Conditional Generation-based Counterfactual Explanations for Multivariate Time Series Classification. Multivariate time series (MTS) are prevalent but inherently complex, making them challenging to analyze due to strong temporal and inter-variable correlations. This complexity often results in the use of sophisticated and difficult-to-interpret machine learning models. In real-life scenarios where critical applications of these models are common, their acceptability is crucial. Counterfactual explanations have emerged as a valuable tool for understanding machine learning systems by providing post-hoc analyses of classification models. We introduce CFE4MTS (CounterFactual Explanation for Multivariate Time Series) [35], a conditional, generation-based, plausible counterfactual explanation method, specifically designed for multivariate time series classification. Our approach leverages advanced time series modeling techniques to generate interpretable counterfactuals that belong to a given target class distribution. To evaluate the effectiveness of our method, we apply it to various real datasets, demonstrating the superiority of our approach over the state of the art methods.

The Potential of Cognitive Circles to Measure Mental Load. In Human-Computer Interaction, Usability, and Interaction Design, obtaining objective measures of mental workload is desirable yet challenging, as

current methods are either costly and intrusive or subjective and unreliable. To overcome these limitations, we devised Cognitive Circles, a technique that estimates workload by analyzing the kinematic properties of circular traces drawn on a tablet as people simultaneously perform cognitively demanding tasks of different types (arithmetic, reading, and spatial reasoning). We investigate the feasibility of this approach and lay the foundations for establishing its viability through a controlled experiment that addresses two questions: (A) Do participants' traces reliably encode information to predict the tasks' difficulty? and (B) Do predictive patterns generalize across tasks in different cognitive activities? Our results show that Cognitive Circles can predict task difficulty with an average accuracy of 75% (reaching up to 94% for spatial reasoning tasks), capturing meaningful signatures of mental workload (A). Prediction performance, however, varies substantially across task types (B), suggesting that each task domain induces people to exhibit distinct kinematic patterns. These findings highlight Cognitive Circles as a promising low-cost approach to workload assessment and point to its potential for informing adaptive HCI and the design of cognitively aware systems [34].

8.3.2 Privacy and Machine Learning

Cross-table Synthetic Tabular Data Detection. Detecting synthetic tabular data is essential to prevent the distribution of false or manipulated datasets that could compromise data-driven decision-making. This study explores whether synthetic tabular data can be reliably identified "in the wild"—meaning across different generators, domains, and table formats. This challenge is unique to tabular data, where structures (such as number of columns, data types, and formats) can vary widely from one table to another. We propose three cross-table baseline detectors and four distinct evaluation protocols, each corresponding to a different level of "wildness". Our very preliminary results confirm that cross-table adaptation is a challenging task [36].

Low-Cost Privacy-Preserving Decentralized Learning. Decentralized learning (DL) is an emerging paradigm of collaborative machine learning that enables nodes in a network to train models collectively without sharing their raw data or relying on a central server. This paper introduces Zip-DL, a privacy-aware DL algorithm that leverages correlated noise to achieve robust privacy against local adversaries while ensuring efficient convergence at low communication costs. By progressively neutralizing the noise added during distributed averaging, Zip-DL combines strong privacy guarantees with high model accuracy. Its design requires only one communication round per gradient descent iteration, significantly reducing communication overhead compared to competitors. We establish theoretical bounds on both convergence speed and privacy guarantees. Moreover, extensive experiments demonstrating Zip-DL's practical applicability make it outperform state-of-the-art methods in the accuracy vs. vulnerability trade-off. Specifically, Zip-DL (i) reduces membership-inference attack success rates by up to 35% compared to baseline DL, (ii) decreases attack efficacy by up to 13% compared to competitors offering similar utility, and (iii) achieves up to 59% higher accuracy to completely nullify a basic attack scenario, compared to a state-of-the-art privacy-preserving approach under the same threat model. These results position Zip-DL as a practical and efficient solution for privacy-preserving decentralized learning in real-world applications [28].

The Privacy Cost of Fine-Grained Electrical Consumption Data. The collection of electrical consumption time series through smart meters grows with ambitious nationwide smart grid programs. This data is both highly sensitive and highly valuable: strong laws about personal data protect it while laws about open data aim at making it public after a privacy-preserving data publishing process. In this work, we study the uniqueness of large scale real-life fine-grained electrical consumption time-series and show its link to privacy threats. Our results show a worryingly high uniqueness rate in such datasets. In particular, we show that knowing 5 consecutive electric measures allows to re-identify on average more than 90% of households in our 2.5M half-hourly electric time series dataset. Moreover, uniqueness remains high even when data is severely degraded. For example, when data is rounded to the nearest 100 watts, knowing 7 consecutive electric measures allows to re-identify on average more than 40% of the households (same dataset). We also study the relationship between uniqueness and entropy, uniqueness and electric consumption, and electric consumption and temperatures, showing their strong correlation [26].

8.3.3 Other Applications

Mitigating analytical variability in fMRI results with style transfer. 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 [30].

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

- **Stellantis - Univ. Rennes (2023-2026) with MALT Team**

Participants: Laurence Rozé, Paul Sevellec.

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 Sevellec (Thèse CIFRE).

- **Orange - Inria with AIstroSight Team (2024-2027)**

Participants: Tassadit Bouadi, Ismail Bachchar.

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

Context. This project is conducted in collaboration with Orange Labs Lannion and focuses on the development of interpretable machine learning methods for high-stakes decision-making systems. As machine learning models are increasingly deployed in industrial applications such as credit acceptance or attribution prediction, ensuring transparency and reliability has become a central methodological challenge. Counterfactual (CF) explanations constitute a widely adopted approach in eXplainable Artificial Intelligence (XAI), as they provide instance-level explanations by identifying minimal feature modifications required to change a model's prediction toward a specified target outcome.

Despite their effectiveness, existing counterfactual methods often lack robustness when confronted with distributional shifts between training and deployment data. This work specifically addresses the

problem of counterfactual robustness under distribution mismatch, a setting that frequently arises in real-world industrial pipelines, where models trained in one context may be deployed in heterogeneous environments. The methodological objective is to design counterfactual generation techniques that remain valid, realistic, and actionable across varying data distributions.

In line with Orange’s commitment to the responsible use of artificial intelligence, this research emphasizes algorithmic transparency and explainability as key enablers of trustworthy AI and its adoption by both end-users and client managers. The CIFRE PhD of Ismail Bachchar, funded by Orange, is dedicated to the development of generic, robust, and industrially applicable counterfactual explanation methods that meet these requirements.

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 on time series classification. In collaboration with Luis Galárraga, Rodne Quijije and Paul Viillard, Gonzalo is working on the development of fast, accurate, and user-friendly feature-attribution and concept-based explanations for convolution-based time series classification. This research avenue emerged from his work on cognitive load estimation using convolution-based time series classifiers, and has the potential to unlock progress of our understanding on state-of-the-art classification and how this enables for accurate cognitive load prediction from circular traces.

10.2 National initiatives

- **#DigitAg: Digital Agriculture**

Participants: Alexandre Termier, Véronique Masson, Christine Largouët, Luis Galárraga, Pierre Cottais.

#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 Alexandre 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, Marine Hamon, Sacha Germain, Christine Largouet, Véronique Masson, Louis Bonneau De Beaufort, Tassadit Bouadi.

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.

- **PEPR IA ADAPTING** with MALT Team

Participants: Luis Galárraga, Julianne Guerbette, Laurence Rozé, Éliisa Fromont.

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 algorithm 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, which general goal is to improve the development of AI software in research teams of CNRS labs.

10.2.1 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.

Period: 03/02/2020 - 31/12/2024 (final scientific activities still going on in 2025)

Budget: 188k€ (Inria)

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 - 30/06/2026

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 Lamerçerie.

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.

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

Participants: Peggy Cellier, Alexandre Termier.

Period: 01/01/2025 – 31/12/2028

Budget: 542k€ (INSA Rennes)

The recent major advances in Artificial Intelligence are to a very large part due to the significant progress in Machine Learning on the topic of Deep Neural Networks, which have been shown to be able to achieve state-of-the-art performance in just about any application area. Such networks have a large number of parameters that interact in intricate ways, which gives them the power to learn complicated concepts but also makes them very difficult to interpret and explain, which strongly limits their applicability in practice, such as in health care. Explainability of graph neural networks (GNN) has recently attracted a lot of research attention.

Existing work mostly focuses on explaining individual neurons, or on learning interpretable input/output mappings, rather than actually explaining what is going on inside the network. In Pandora, our

hypothesis is that a GNN performs well because it has been able to learn important concepts within the data. These concepts deserve to be brought to the attention of experts to develop new scientific breakthroughs or to detect biases within the training data. Our research hypothesis is that we can provide knowledge by introspecting the GNN models. With Pandora, we propose to characterize, gain insight, and explain in easily understandable terms the inner workings of GNNs. In a nutshell, we propose to discover statistically significant patterns of neural co-activation so as to determine how networks encode concepts over multiple neurons, identify information shared between classes, trace information through the network, and overall, to determine how networks perceive the world. Using those patterns we want to characterise under which conditions a prediction made by the network is to be trusted, and finally, learn trustworthy GNNs that are explicitly explainable using patterns. To assess the usefulness of our work, we will apply it on a variety of use cases in chemoinformatics, social web and semantic web.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

Member of the organizing committees.

- Organization and Chairing of the **AIMLAI** Workshop (Advances in Interpretable Machine Learning and Artificial Intelligence) at ECML/PKDD (**Luis Galárraga**, **Tassadit Bouadi**)
- **Peggy Cellier** is in charge with Marie Tahon (Université du Mans) of the "comité de pilotage" collègue TLH (Technologies du Langage Humain) of AFIA (Association française pour l'Intelligence Artificielle) since 2025 (and member since 2024).

11.1.2 Scientific events: selection

Chair of conference program committees

- **Peggy Cellier** was a program chair of the 2nd International Joint Conference on Conceptual Knowledge Structures (CONCEPTS) in September 2025, Cluj-Napoca. This conference merges three existing international conferences: ICCS, ICFCA, and CLA.

Member of the conference program committees.

- **Peggy Cellier**: Senior PC of **ECMLPKDD'25**, Senior PC of **EGC'25**, Senior PC of **IDA'25**
- **Alexandre Termier**: Area Chair of **ICDM'25**, PC for **ECMLPKDD'25**, PC for **AIMLAI'25 workshop**
- **Tassadit Bouadi**: PC of **IDA'25**, PC of **ECAI'25**
- **Christine Largouët**: PC of **ECAI'25**, PC of **PFIA APIA'25**,
- **Laurence Rozé**: PC of **ECAI'25**

Reviewer. The Web Conference (Luis Galárraga), ISWC (Luis Galárraga), XAI (Luis Galárraga), IJCNN (Luis Galárraga), IDA (Sébastien Ferré), CONCEPTS (Sébastien Ferré), FQAS (Sébastien Ferré)

11.1.3 Journal

Member of the editorial boards

- **Alexandre Termier**: Editorial Board of Data Mining and Knowledge Discovery

Reviewer - reviewing activities.

- **Peggy Cellier**: Data Mining and Knowledge Discovery
- **Alexandre Termier**: Data Mining and Knowledge Discovery
- **Luis Galárraga**: Data and Knowledge Engineering
- **Sébastien Ferré**: International Journal of Advanced Research
- **Christine Largouët**: AAPG ANR projects.

11.1.4 Invited talks

- Keynote talk at RJCIA (Rencontres des Jeunes Chercheurs en Intelligence Artificielle), part of the PFIA platform for scientific events about AI in France (Luis Galárraga, July 2025)
- Invited talk and “Grand Témoin” at the **Assises de la Recherche et de l’Innovation en Côtes d’Armor**, topic: introduction to generative AI, Alexandre Termier, November 2025
- Invited talk at the “Ecole Chercheur : Données et Modèles” of INRAE, topic: introduction to generative AI, Alexandre Termier, October 2025
- **Panel at BDA’25** “Les BDs pourront-elles sauver l’IA”, Alexandre Termier, October 2025.
- Presentation of the WAIT4 project at **Breizh Carnot Tech**, Alexandre Termier, November 2025.
- Presentation at the **SPACE Rennes (salon de l’élevage)**, “État de l’art de l’IA et enjeux pour l’agriculture et l’élevage”, Alexandre Termier, September 2025.
- Introductory talk at the “Semaine de l’IA”, Univ. Rennes - ISTIC, Alexandre Termier, September 2025

11.1.5 Leadership within the scientific community

- **Peggy Cellier** was 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.

11.1.6 Scientific expertise

- **Alexandre Termier**: MIAI Cluster Chair project
- **Tassadit Bouadi**: Member of the working group of Axis 2 ‘Research and Innovation Program’ within the IRIS-E program at the University of Rennes
- **Christine Largouët**: Member of the CSTP, PEPR Agroecology and ICT; Member of the CS, INRAE PHASE department.

11.1.7 Research administration

- **Peggy Cellier** was in charge of the Phd students of the IRISA lab (commission personnel each month, etc). She is also a member of "Conseil de l’école doctorale MATISSE". Both until the mid-2025. Since September 2025, she serves in the "Commission personnel" of IRISA/Inria RBA for the visitor applications.

11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

Apart from Luis Galárraga (research scientist), 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

- **Luis Galárraga** is in charge of the module Knowledge Representation and Semantic Web (RPCO) at the M1 IA offered by the University of Rennes (Feb - Apr 2025, 16.5h of CM), assisted by Isseinie Sinouassane (19h of TD and TP). He also taught 3h within the course “Data Mining and Visualization” (by Alexandre Termier M2 SIF, Nov 2025).
- **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).
- **Sébastien Ferré** was the head of Master M1 Miage, and of the EIT international master track in Data Science (about 75 students), until July.
- **Peggy Cellier** is the head of the last year at Computer Science Department at INSA (master 2 level, about 70 students).
- **Tassadit Bouadi** Since September 2023, she has been co-head, together with Romaric Gaudel, of the Master’s program in Artificial Intelligence (Master 1 and Master 2) at ISTIC, University of Rennes, which they jointly created and implemented, and for which they are responsible for academic coordination and strategic development. She is also responsible for the work-study program of the Master 1 AI.
- **Christine Largouet** is co-head of the master M1 and M2 E2C (Water, Energy and Climate, climate change mitigation and adaptation) at Institut Agro Rennes Angers. She was head of the computer science educational unit at Institut Agro Rennes Angers (2 engineering schools) from septembre 2006 until septembre 2024.
- **Laurence Rozé** is the head of the L2 studies at INSA of Rennes (296 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), Data Mining (M2 IAA, co-head with Nathalie Girard).
- **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.
- **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). At master 2 SIF, she teaches in English 4,5 hours in the data mining course (DMV). She also teaches at University of Rennes, Lienc 1 BioMIA: Introduction to AI.
- **Sébastien Ferré** was responsible of 4 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), "Technological Watch" (M1 Miage EIT).
- **Romaric 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 ISTIC (Univ Rennes) : "Algorithmique pour l'IA" (Master 1 IA), option "IA et Jeux" (Master 1 IL), "Réussir son insertion professionnelle" (Master 1 IA), and "Bases de données" (L2 Informatique).

- **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), ADS (L2).
- **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 also represents INSA Rennes in the CMA (Compétence et Métier d'Avenir) IA TIARe and Cluster IA SequoIA.
- **Alexandre Termier** is an elected member of the Department Committee (conseil d'UFR) of the ISTIC department 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, ...)

11.2.1 Supervision

Internships

- Baptiste Amice (M2, Feb 2025 - July 2025, supervised by Peggy Cellier and Sébastien Ferré) with the subject "LLM pour l'interrogation de données du web sémantique".
- Lydia Achour (L3, May 2025 - Aug. 2025, supervised by Luis Galárraga, Christine Largouët, Gonzalo Méndez) with the subject "Scrollytelling Explanations for AI systems".
- Isidore Gomendy (L3, June 2025 - July 2025, supervised by Luis Galárraga and Peggy Cellier) with the subject "Embedding Rules with Knowledge Graph Embeddings".

PhD Students

- Ismail Bachchar (2024-2027, OrangeLabs, supervised by Tassadit Bouadi, Thomas Guyet from AISTROSIGHT team, and Françoise Fessant from Orange) with the subject "Generation of stable and robust explanations."
- Vanessa Fokou (2022-2025, supervised by Florence Le Ber, Xavier Dolques from Univ. Strasbourg and Peggy Cellier, Sebastien Ferre) with the subject "Comparison and cooperation of different Formal Concept Analysis approaches for relational data"
- Sacha Germain (2024-2027, Inria, supervised by Tassadit Bouadi, Christine Largouet, Laurence Rozé) with the subject "Detection and explanation of individual and collective behavior within a group to assess their well-being"
- Julianne Guerbette (2025-2028, Univ. Rennes, supervised by Luis Galárraga, Laurence Rozé) with MALT Team on the subject "Continual Neuro-symbolic Learning of Knowledge Graph Embeddings", financed by the PEPR project AdaptING.
- Gwladys Kelodjou (2022-2026, supervised by Véronique Masson, Laurence Rozé, Alexandre Termier), with the subject "Beyond Divination: Stabilizing the Interpretability of Machine Learning Algorithms"
- Isseïnie Sinouvassane (2023-2026, ENS Rennes, supervised by Alexandre Termier, Luis Galárraga) on the subject "How-Provenance Polynomials for Efficient and Greener Rule Mining", financed by an ENS doctoral scholarship.

- Paul Sevellec (2024-2027, Univ. Rennes , supervised by Elisa Fromont, Romaric Gaudel, Laurence Rozé) on the subject “Explications de séries temporelles multivariées par contrefactuels”.

Engineers

- Frederic Lang, 2024-2026; supervised by Sebastien Ferre, 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.
- Pierre Cottais, 2025-2026; supervised by Alexandre Termier, Peggy Cellier; project: DigitAg. He works on the analysis of precision farming data in collaboration with INRAE colleagues from Pegase. Currently our focus is heat stress data from dairy cows equipped with sensors (accelerometers and temperature).
- Marine Hamon, 2025-2026; supervised by Alexandre Termier, Peggy Cellier; project: WAIT4. She works on the analysis of precision farming data in collaboration with INRAE colleagues from Pegase. Currently our focus is heat stress data from dairy cows equipped with sensors (accelerometers and temperature).

Postdoctoral students

- Aurélien Lamercherie, 2024-2026; supervised by Sebastien Ferre and Peggy Cellier; project: MEKANO.

11.2.2 Juries

PhD Juries.

- **Peggy Cellier** was a member of the following PhD juries in 2025: Marion Schaeffer, 28/03 INSA Rouen (reviewer); Lucas Potin, 02/09 Avignon Université (reviewer).
- **Luis Galárraga** was examiner in the PhD juries of Sacha Corbugy (29/09/2025 University of Namur) and Ataollah Kamal (01/09/2025, INSA Lyon)
- **Sebastien Ferre** was a member of the following PhD juries in 2025: Aymen Bazouzi, 26/01 Univ. Rennes (president); Sarra Ouelhadj, 21/01 Univ. Lyon 1 (rapporteur); Ginwa Fakh, 12/09 Nantes Univ. (rapporteur).
- **Alexandre Termier** was a member of the following juries: Josha Cüppers, PhD, Saarland University, 11/9 (reviewer) ; Erwan Vincent, PhD, Univ. Rennes, 4/12 (president).
- **Christine Largouët** was a member of the PHD jury of Loïc Eyango, Université of Nantes, 04/06 (reviewer).

11.2.3 Doctoral advisory committee (CSID)

- **Peggy Cellier** was a member of the mid-term evaluation committee of Clémence Sebe (Université Paris Saclay); Randa Bendjeddou (Université de Lyon 2); Yacine Mokhtari (IMT Atlantique Brest).
- **Tassadit Bouadi** was a member of the mid-term evaluation committee of Estelle Yvana Eyenga Abate (INSA Rennes).

11.2.4 Educational and pedagogical outreach

- Introductory Talk to AI in the workshop “Manipuler l’intelligence artificielle pour l’enseigner” organized by “Maison de la Science” at Univ. Rennes (Luis Galárraga, Jan 2025).

11.3 Popularization

11.3.1 Participation in Live events

- Chair of the dissemination workshop on “Introduction to AI” organized by L’Atelier du 5 Bis of Dinan (April 2025)
- Invited speaker to the dissemination conferences (Café de l’IA) “L’Intelligence Artificielle et Nous” and “Applications Positives de l’IA” organized by the Médiathèque de Dinard on January 24 and 29, 2025
- **Peggy Cellier** was invited to give introduction talk (in French) for three events : Lions Club of Angers meeting, Club des quarantième of Chollet and "Journée de l’IA en santé" at UFR Médecine of Rennes. She also participated to the event "Elles bougent pour l’orientation" at Collège Jacques Brel Noyal-sur-Vilaine.
- **Tassadit Bouadi** was invited to give AI introduction talks (in French) for two events : Journées Parité de la communauté mathématiques and Journées "Filles, Maths et Informatique, une équation lumineuse".

11.3.2 Others science outreach relevant activities

- Animation of the different Inria dissemination stands set for the “Semaine de la Science” at Champs Libres (04/10/2025, Luis Galárraga, Isseinie Sinouvassane, Sacha Germain)
- **Tassadit Bouadi** is co-organizer of the project L Codent, L Créent Rennes, since 2018.

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. Bouadi and T. Guyet. ‘Generating Efficiently Realistic Counterfactual Explanations’. In: *Machine Learning* 115.2 (23rd Jan. 2026), p. 27. doi: [10.1007/s10994-025-06947-2](https://doi.org/10.1007/s10994-025-06947-2). URL: <https://inria.hal.science/hal-05472901>.
- [17] 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>.
- [18] 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>.
- [19] 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>.

- [20] 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

- [21] L. Bonneau de Beaufort, C. Xavier, D. Renaudeau, C. Largouët and F. Gondret. ‘Data Paper: HotPig, a behavioural dataset of pigs under heat stress’. In: *Animal - Open Space* 5 (2026), p. 100112. doi: [10.1016/j.anopes.2025.100112](https://doi.org/10.1016/j.anopes.2025.100112). URL: <https://hal.inrae.fr/hal-05410799> (cit. on p. 14).
- [22] J. Delaunay, L. Galárraga, C. Largouët and N. V. Berkel. ‘Impact of Explanation Technique and Representation on Users’ Comprehension and Confidence in Explainable AI’. In: *Proceedings of the ACM on Human-Computer Interaction* 9.2 (Apr. 2025), Article CSCW113. doi: [10.1145/3711011](https://doi.org/10.1145/3711011). URL: <https://hal.science/hal-04948723> (cit. on p. 16).
- [23] V. Fokou, P. Cellier, X. Dolques, S. Ferré and F. Le Ber. ‘Theoretical comparison of Relational Concept Analysis (RCA) and Graph-FCA (GCA)’. In: *International Journal of Approximate Reasoning* 186 (Nov. 2025), p. 109496. doi: [10.1016/j.ijar.2025.109496](https://doi.org/10.1016/j.ijar.2025.109496). URL: <https://hal.science/hal-05138994> (cit. on p. 15).
- [24] 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’. In: *PLoS ONE* (2025), pp. 1–39. doi: [10.1371/journal.pone.0317566](https://doi.org/10.1371/journal.pone.0317566). URL: <https://inserm.hal.science/inserm-04465765>.
- [25] V. Guyomard, F. Fessant, T. Bouadi and T. Guyet. ‘Generating Efficiently Realistic Counterfactual Explanations’. In: *Machine Learning* 115.2 (23rd Jan. 2026), p. 27. doi: [10.1007/s10994-025-06947-2](https://doi.org/10.1007/s10994-025-06947-2). URL: <https://inria.hal.science/hal-05472901> (cit. on p. 15).
- [26] A. Voyez, T. Allard, G. Avoine, P. Cauchois, E. Fromont and M. Simonin. ‘The Privacy Cost of Fine-Grained Electrical Consumption Data’. In: *Scientific Reports* 15.17391 (19th May 2025), pp. 1–8. doi: [10.1038/s41598-024-78285-7](https://doi.org/10.1038/s41598-024-78285-7). URL: <https://hal.science/hal-03833605> (cit. on p. 17).

Invited conferences

- [27] F. Gondret, C. Verplanck, R. Lardy, M. Taghipoor, M. Bonneau and A. Termier. ‘De nouveaux outils et algorithmes au service de la mesure du bien-être des animaux face au changement climatique’. In: 2025 - Journées scientifiques du programme PEPR Agroécologie et Numérique. Dijon (FR), France, 2025. URL: <https://hal.inrae.fr/hal-04920261>.

International peer-reviewed conferences

- [28] S. Biswas, D. Frey, R. Gaudel, A.-M. Kermarrec, D. Lerévérénd, R. Pires, R. Sharma and F. Taïani. ‘Low-Cost Privacy-Preserving Decentralized Learning’. In: *Proceedings on Privacy Enhancing Technologies Symposium*. Privacy Enhancing Technologies Symposium. Vol. 2025. 3. Washington DC, United States, July 2025, pp. 451–474. doi: [10.56553/popets-2025-0108](https://doi.org/10.56553/popets-2025-0108). URL: <https://hal.science/hal-04993586> (cit. on p. 17).
- [29] V. Fokou, P. Cellier, X. Dolques, S. Ferré and F. Le Ber. ‘Relational Concept Analysis and Graph-FCA in Practice: Analyzing Multi-relational Data’. In: *Conceptual Knowledge Structures*. CONCEPTS 2025 - Second International Joint Conference on Conceptual Knowledge Structures. Vol. 15941. Lecture Notes in Computer Science. Cluj-Napoca, Romania: Springer Nature Switzerland, 6th Sept. 2025, pp. 35–50. doi: [10.1007/978-3-032-03364-2_3](https://doi.org/10.1007/978-3-032-03364-2_3). URL: <https://univoak.hal.science/hal-05298483>.

- [30] E. Germani, C. Maumet and E. Fromont. ‘Mitigating analytical variability in fMRI results with style transfer’. In: *Medical Imaging with Deep Learning (MIDL 2025)*. Salt Lake City (US), United States, 9th July 2025. URL: <https://inserm.hal.science/inserm-04531405> (cit. on p. 18).
- [31] G. C. N. Kindji, E. Fromont, L. M. Rojas-Barahona and T. Urvoy. ‘Synthetic Tabular Data Detection In the Wild’. In: *IDA 2025 - International Symposium on Intelligent Data Analysis*. Konstanz, Germany, 7th May 2025. DOI: [10.1007/978-3-031-91398-3_7](https://doi.org/10.1007/978-3-031-91398-3_7). URL: <https://hal.science/hal-04970770>.
- [32] A. Lamercurie, P. Cellier and S. Ferré. ‘Web-SPARQL: Hybrid Querying over Knowledge Graphs, Web, and Microdata’. In: *Web Engineering - 25th International Conference, ICWE 2025, Delft, The Netherlands, June 30 - July 3, 2025, Proceedings*. ICWE 2025 - 25th International Conference on Web Engineering. Vol. 15749. Lecture Notes in Computer Science. Delft, Netherlands: Springer Nature Switzerland, 30th June 2025, pp. 193–208. DOI: [10.1007/978-3-031-97207-2_15](https://doi.org/10.1007/978-3-031-97207-2_15). URL: <https://hal.science/hal-05460867> (cit. on p. 15).
- [33] 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>.
- [34] G. G. Méndez, L. Galárraga, R. Quijije and M. A. Nacenta. ‘The Potential of Cognitive Circles to Measure Mental Load’. In: *UIST’25 Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technolog*. UIST 2025 - 38th Annual ACM Symposium on User Interface Software and Technology. Busan, South Korea: ACM, 27th Sept. 2025, pp. 1–10. DOI: [10.1145/3746059.3747716](https://doi.org/10.1145/3746059.3747716). URL: <https://inria.hal.science/hal-05396833> (cit. on p. 17).
- [35] P. Sevellec, E. Fromont, R. Gaudel, L. Rozé and M. Sammarco. ‘Plausible Conditional Generation-based Counterfactual Explanations for Multivariate Times Series Classification’. In: *Proceedings of the 28th European Conference on Artificial Intelligence, IOS Press*. ECAI 2025 - European Conference on Artificial Intelligence. Bologna, Italy, 25th Oct. 2025. URL: <https://hal.science/hal-04928456> (cit. on p. 16).

Conferences without proceedings

- [36] 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 at COLING*. Abu dahbi, United Arab Emirates, 2025, pp. 1–7. URL: <https://hal.science/hal-04840865> (cit. on p. 17).
- [37] Y. Mahé, S. Leplaideur, E. Fromont, E. Bannier and F. Galassi. ‘Détection et segmentation des anomalies cérébrales post-AVC’. In: *IABM 2025 - Colloque Français d’Intelligence Artificielle en Imagerie Biomédicale*. Nice, France, 2025, pp. 1–2. URL: <https://hal.science/hal-04910273>.
- [38] P. Molli, H. Skaf-Molli, S. Ferré, A. Gaignard and P. Cellier. ‘SparqLLM : Retrieval-Augmented SPARQL Query Processing’. In: *ESWC 2025 - 22nd European Semantic Web Conference*. Portoroz, Slovenia, 2025, pp. 1–5. URL: <https://hal.science/hal-05138658> (cit. on p. 15).

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*. Frontiers in Artificial Intelligence and Applications. IOS Press, July 2025, pp. 1–22. DOI: [10.3233/FAIA250224](https://doi.org/10.3233/FAIA250224). URL: <https://hal.science/hal-04792154> (cit. on p. 15).

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

- [40] *Conceptual Knowledge Structures*. CONCEPTS 2025. Vol. 15941. Lecture Notes in Computer Science. Cluj - Napoca, Romania: Springer Nature Switzerland, 2025. DOI: [10.1007/978-3-032-03364-2](https://doi.org/10.1007/978-3-032-03364-2). URL: <https://hal.science/hal-05460872> (cit. on p. 14).

- [41] *Selected papers from the Second International Joint Conference on Conceptual Knowledge Structures* 187 (Dec. 2025). DOI: [10.1016/J.IJAR.2025.109545](https://doi.org/10.1016/J.IJAR.2025.109545). URL: <https://hal.science/hal-05460874>.

Reports & preprints

- [42] G. Kelodjou, L. Rozé, V. Masson, R. Gaudel, L. Galárraga and A. Termier. *Stratum by Stratum: Building Stable SHAP Explanations through Layered Approximations*. 23rd Feb. 2025. URL: <https://hal.science/hal-04962944> (cit. on p. 16).

12.3 Cited publications

- [43] S. Borzsony, D. Kossmann and K. Stocker. ‘The Skyline operator’. In: *Proceedings 17th International Conference on Data Engineering*. 2001, pp. 421–430. DOI: [10.1109/ICDE.2001.914855](https://doi.org/10.1109/ICDE.2001.914855) (cit. on p. 8).
- [44] P. D. Grünwald. *The Minimum Description Length Principle*. The MIT Press, Mar. 2007. DOI: [10.7551/mitpress/4643.001.0001](https://doi.org/10.7551/mitpress/4643.001.0001). URL: <https://doi.org/10.7551/mitpress/4643.001.0001> (cit. on p. 8).
- [45] S. M. Lundberg and S.-I. Lee. ‘A unified approach to interpreting model predictions’. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Long Beach, California, USA, 2017, pp. 4768–4777 (cit. on p. 8).
- [46] M. T. Ribeiro, S. Singh and C. Guestrin. ‘“Why Should I Trust You?”: Explaining the Predictions of Any Classifier’. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. Ed. by B. Krishnapuram, M. Shah, A. J. Smola, C. C. Aggarwal, D. Shen and R. Rastogi. ACM, 2016, pp. 1135–1144. DOI: [10.1145/2939672.2939778](https://doi.org/10.1145/2939672.2939778). URL: <https://doi.org/10.1145/2939672.2939778> (cit. on p. 8).