



HAL
open science

2023 Activity Report – Orpailleur Team (LORIA) : Knowledge Discovery and Knowledge Engineering

Amedeo Napoli, Alexandre Blanché, Lydia Boudjeloud-Assala, Briec Conan-Guez, Miguel Couceiro, Alain Gély, Frédéric Pennerath, Yannick Toussaint, Mario Valencia-Pabon

► To cite this version:

Amedeo Napoli, Alexandre Blanché, Lydia Boudjeloud-Assala, Briec Conan-Guez, Miguel Couceiro, et al.. 2023 Activity Report – Orpailleur Team (LORIA): Knowledge Discovery and Knowledge Engineering. Université de Lorraine, CNRS, LORIA. 2024. hal-04575652

HAL Id: hal-04575652

<https://hal.science/hal-04575652>

Submitted on 15 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

2023
ACTIVITY REPORT

Team
Orpailleur

Knowledge Discovery and Knowledge
Engineering

DEPARTMENT

Knowledge and language management

Contents

Team Orpailleur	1
1 Team members, visitors, external collaborators	1
2 Overall objectives	2
3 Research program	3
3.1 Axis 1. Exploratory and hybrid knowledge discovery	3
3.2 Axis 2. Explainability, transparency and fairness	3
3.3 Axis 3. Analogy based reasoning	4
3.4 Axis 4. Computational Complexity	5
4 Application domains	6
4.1 Healthcare	6
4.2 Agronomy	6
4.3 Material science	7
5 Social and environmental responsibility	7
6 New softwares, platforms, open data	8
6.1 New software	8
6.1.1 FixOut	8
6.1.2 ANNa	8
6.1.3 Caspailleur and Paspailleur	9
6.2 New platforms	9
6.2.1 AT2TA-Nexus	9
6.3 Open data	10
7 New results	10
7.1 Axis 1. Exploratory and hybrid knowledge discovery	10
7.1.1 Connexions between FCA and Decision Trees (description quivers)	10
7.1.2 A study of the closure algorithms and the discovery of implication bases	10
7.1.3 A study of data complexity from an FCA point of view	10
7.2 Axis 2. Explainability, transparency and algorithmic fairness	11
7.2.1 Fairness Notions and Related Tensions	11
7.2.2 Statistical approaches to detect disparity prone features in group fairness settings	11
7.2.3 Clarity	12
7.2.4 Bayesian Deep Learning and Reference Priors	12
7.3 Axis 3. Analogy based reasoning	12
7.3.1 Galois theory for analogical classifiers	12
7.3.2 Case-base reasoning and analogical transfer.	13
7.3.3 Knowledge Graph Bootstrapping via Zero-Shot Analogical Pruning	13
7.4 Axis 4. Computational complexity	13
7.4.1 Fine-grained and parameterized complexity of graph coloring problems.	13
7.4.2 Rotation distance of split graph associahedra is polynomial	14
7.4.3 Exact distance Kneser graphs	14
8 Bilateral contracts and grants with industry	15
8.1 Bilateral contracts with industry	15
8.1.1 Happiso	15
8.1.2 Orange	15
8.2 Grants with industry	15
8.2.1 Kesaio (2022–2024)	15
8.3 Transfer with Industry	15
8.3.1 FixOut start-up	15

9 Partnerships and cooperations	16
9.1 International research visitors	16
9.1.1 Visits of international scientists	16
9.1.2 Visits to international teams	16
9.2 European initiatives	16
9.3 National initiatives	17
9.3.1 ANR	17
10 Dissemination	21
10.1 Promoting scientific activities	21
10.1.1 Scientific events: organisation	21
10.1.2 Scientific events: selection	21
10.1.3 Journal	21
10.1.4 Invited talks	22
10.1.5 Leadership within the scientific community	22
10.1.6 Scientific expertise	22
10.2 Pedagogical responsibilities - Supervision - Juries	22
10.2.1 Pedagogical responsibilities	22
10.2.2 Supervision	23
10.2.3 Juries	23
10.3 Popularization	24
10.3.1 Interventions	24
10.3.2 Education	24
11 Scientific production	24
11.1 Major publications	24
11.2 Publications of the year	24
11.3 Cited publications	27

Team Orpailleur

Keywords

Knowledge discovery in databases, pattern mining, formal concept analysis, redescription mining, knowledge engineering, explainability, algorithmic fairness, analogy based and case-based reasoning, algorithmic complexity.

1 Team members, visitors, external collaborators

Research Scientists

- Amedeo Napoli [CNRS (Emeritus), Researcher]

Faculty Members

- Alexandre Blansché [UL, Associate Professor]
- Lydia Boudjeloud-Assala [UL, Associate Professor, HDR]
- Brieuc Conan-Guez [UL, Associate Professor]
- Miguel Couceiro [UL, Team leader, Professor]
- Alain Gély [UL, Associate Professor]
- Frédéric Pennerath [CentraleSupélec, Associate Professor]
- Yannick Toussaint [UL, Professor, Director of LORIA since 2024]
- Mario Valencia [UL, Associate Professor, HDR]

Post-Doctoral Fellows

- Simon Vilmin [CNRS, Post-Doctoral Fellow, until Sep. 2023]

PhD Students

- Nacira Abbas [UL, PhD Student, Defended Oct. 2023]
- Ambroise Baril [UL, PhD Student, since Sep. 2021]
- Latifa Bouchekif [UL, PhD Student, since Sep. 2023]
- Egor Dudyrev [UL, PhD Student, since Jan. 2023]
- Mohamed Fellaji [UL, PhD Student, since Jan. 2022]
- Félix Gaschi [Posos-UL, PhD Student, Cifre, Defended Dec. 2023]
- Lucas Jarnac [Orange-UL, PhD Student, Cifre, since Feb. 2023]
- Esteban Marquer [UL, PhD Student, since Oct. 2020]
- Morgan Ruiz-Huidobro [Happiso-UL, PhD Student, Cifre, since Apr. 2023]
- Claire Theobald [UL, PhD Student, Defended Jun. 2023]
- Laura Zanella [UL, PhD Student, Defended Dec. 2023]
- Georgios Zervakis [Inria, PhD Student, Defended Mar. 2023]

Technical Staff

- Cherif-Hassan Noursradine [UL, Engineer, since Oct. 2023]
- Mamadou Ndiaye [CNRS, Engineer, since Nov. 2022]

Interns and Apprentices

- Hamza Azizi [UL, Intern, June-November 2023]

Administrative Assistants

- Delphine Hubert [UL]
- Anne-Marie Messaoudi [CNRS]

External Collaborators

- Guilherme Alves [FixOut CEO]
- Jaume Baixeries [Universitat Politecnica de Catalunya]
- Víctor Codocedo [Universidad Técnica Federico Santa María]
- Mehdi Kaytoue [Infologic Lyon, HDR]
- Sergei O. Kuznetsov [National Research University Higher School of Economics, HDR]
- Victor Langerkvist [Linköping University, HDR]
- David Leake [Indiana University, US, HDR]
- Erkkö Lehtonen [Khalifa University, HDR]
- Yves Lepage [University of Waseda, HDR]
- Pierre-Alexandre Murena [Technische Universität Hamburg]

2 Overall objectives

The research topics of the Orpailleur Team of 2023 gravitated the federating theme “Exploratory and hybrid knowledge discovery” of the team, but opened to new research directions, such as, “Explainability, transparency and fairness” and “Analogy based reasoning” that bridge the gap between Knowledge Representation and Reasoning (KRR) and in Machine Learning (ML). The Orpailleur Team has always advocated for “Knowledge Discovery guided by Domain Knowledge”, a research line which gained a role of paramount importance in ML and Artificial Intelligence (AI). In particular, the recent seminal paper on “Interpretable Machine Learning” [42] lists 10 grand challenges in ML. In these challenges, it can be noticed that knowledge integration in ML is a basic ingredient quite everywhere, while emerging topics are materialized by analogy based reasoning, explainable and trustworthy ML&AI models, in particular, topics pertaining to algorithmic complexity and fairness. These are all represented in the Orpailleur’s research program of 2023.

3 Research program

3.1 Axis 1. Exploratory and hybrid knowledge discovery

In this first research line, we are interested in designing “Hybrid KD Systems” based on symbolic and subsymbolic approaches. Such hybrid systems will combine various components depending on the openness and supervision of the KD algorithms. For example, some components may be associated with black-box ML models such as neural networks with more open ML models such as pattern mining algorithms and decision trees. As in every complex problem-solving task, there is no universal approach which may be used alone to tackle and to solve every problem. Hence, a reasonable strategy is to design a hybrid KD process based on several tactics, following “ensemble learning” principles. Building such an architecture is an open problem and a possible direction is illustrated by “Exploratory Knowledge Discovery” as promoted in the Orpailleur Team3. In Exploratory KD we consider four main dimensions related to data, knowledge, problem-solving, and interactivity. The “data dimension” is connected to the input of KD and involves data preparation, including feature selection, dimensionality reduction, and data transformation. The exploration of the data and pattern spaces are first order operations. The “knowledge dimension” is connected to the data domain and depends on domain knowledge expressed in ontologies and knowledge graphs, or as constraints and preferences. Moreover, a particular attention should be given to the output of KD, where the discovered patterns should be effectively represented as actionable knowledge units. The “problem-solving dimension” is related to the tasks to be achieved and is dependent on iteration and search strategies. The latter can be data-directed or goal-directed, and rely on declarative or procedural approaches. Finally “interactivity” involves an analyst who may integrate constraints and preferences for guiding and minimizing data exploration, and as well assess pattern quality. More practically, in Exploratory KD data seeds or initial patterns may be identified for guiding the pattern and data space exploration, that should be consistent w.r.t. domain knowledge and associated constraints. Threshold issues w.r.t. analyst queries, such as minimal support or confidence, can be addressed thanks to integration of domain knowledge or preferences. Exploration involves interaction, iteration, and the replay of the KD process. It can be based on data-directed or pattern-directed methods, depends on specific interestingness measures, and be guided thanks to visualization procedures. Moreover, exploratory KD should be automated as much as possible in avoiding tedious parameter settings, while keeping facilities for interaction and iteration.

3.2 Axis 2. Explainability, transparency and fairness

Explainability and fairness emerged as an important requirement to guarantee that machine learning (ML) based decision systems can be safely used in practice. Using such systems while explanations are not provided and fairness is not guaranteed, can lead to both distrust in knowledge discovery (KD) and unfair algorithmic decision making (ADM). For instance, in the form of unintended bias typically discriminating against disadvantaged populations such as racial minorities, women, poverty stricken districts, etc. With this recent interest, a multitude of explanation methods and fairness notions have been proposed to capture different aspects of the subjective human reasoning and decisions. Interestingly, although the different notions of algorithmic fairness appear internally consistent, several of them cannot hold simultaneously and hence are mutually incompatible [5]. As a consequence, practitioners assessing and/or implementing explanation methods and fairness requirements need to choose among them. In addition, there are tensions between fairness notions and other desirable properties of ML algorithms, and further research is being pursued to find a good trade-off given the use case at hand.

Fairness through unawareness is one of the simplest approaches to address fairness and consists in dropping the sensitive features before training the ML model. Such approach is an example of process fairness that focuses on the fairness of the learning process rather than the fairness of the output. In this line of research, we recently proposed a framework to handle the fairness/accuracy trade-off ML and KD algorithms based on the idea of *feature dropout* followed by an *ensemble approach* to improve model fairness. All these ideas have been implemented in FixOut, a human-centered and model-agnostic framework and adaptable to both tabular and textual data [31]. Given a classifier, a dataset and a set of sensitive features, FixOut first assesses whether the classifier is fair by checking its reliance on sensitive features using explanations. If deemed unfair, FixOut then applies feature dropout to obtain a pool of

classifiers which are combined into an ensemble classifier that will be less dependent on sensitive features without compromising the whole classifier accuracy. It is also worth mentioning that FixOut is currently under maturation for a startup creation. Moreover, we are also developing specific debiasing solutions for large language models (such as Bert and GPT like models) in which we explore reinforcement learning techniques in a multi-objective setting for trade-off management [29].

Another research direction aimed at achieving greater fairness in the behaviour of deep networks is to make them capable of doubting about their own decisions. Whereas conventional approaches can only predict the component of uncertainty intrinsic to the data (known as aleatoric uncertainty), *Bayesian deep learning* can uniquely measure a second component of uncertainty, known as *epistemic uncertainty* that is linked to the lack of necessary information. While the theory is appealing, the calibration of the various uncertainty scores produced by Bayesian neural networks remains imperfect in practice, making this problem a subject of active research. For instance, in the framework of Claire's thesis [21], we focused on the problem of calibrating epistemic uncertainty in the context of regression problems applied to cosmology.

3.3 Axis 3. Analogy based reasoning

Following the tracks of “Hybrid Knowledge Discovery”, we aim to bridge the two basic concerns in Artificial Intelligence (AI), namely, “Reasoning” and “Learning”. Indeed, over the past years Knowledge Representation and Reasoning (KRR) and Machine Learning (ML) saw a considerable progress but somehow specialized themselves in separate lines of research, thus creating an apparent gap between KRR and ML. To some extent, KRR deals with knowledge and privileges symbolic and discrete approaches, whereas ML handles data and focuses on subsymbolic methodologies. Despite justifiable, this dichotomic view is not accurate, as KRR should deal with data (e.g., through Formal Concept Analysis and Pattern Mining), in the same way that ML incorporates symbolic knowledge (e.g., through Inductive Logic Programming and Neuro-Symbolic Reasoning), and may rely on uncertainty representations (as in Bayesian networks). Bridging the gap between KRR and ML will unravel their potential towards transparency and explainability, e.g., through inference and causal understanding, such as in rule and utility based approaches or in analogical reasoning (AR) and transfer learning.

Analogical reasoning establishes parallels between two situations, which are similar in some respects and dissimilar in others. If such a parallel holds to some extent, there are pairs of items (a, b) and (c, d) such as “ a is to b in situation 1 as c is to d in situation 2”. In particular, statements of the form “ a is to b as c is to d ” are called an analogical proportions. Items a, b, c, d are supposed to be described by a set of related attributes and represented by tuples of attribute values. Attributes may be Boolean, nominal (i.e., finite attribute domains), or even structured as in Natural Language Processing (NLP). Also, analogical inference relies on the idea that if four items a, b, c, d are in analogical proportion for each of the n attributes describing them, it may still be the case for another attribute. For instance, if class labels are known for a, b, c and unknown for x , then one may infer the label for x as a solution of an analogical equation $a : b :: c : x$. This idea has been successfully applied in different inference tasks including case-based reasoning, classification and recommendation problems as well as enlarging training sets for learning classifiers, especially in environments with few labeled examples. For instance, we have recently established a Galois theory of analogical classifiers that put in correspondence formal models of analogy with classifiers compatible with the analogical inference principle [6].

But analogical reasoning can also be viewed and performed at a meta level, through the idea of transfer learning where the idea is to take advantage of what has been learned on a source domain in order to improve the learning process in a target domain related to the source domain. When studying a new problem or a new domain, it is natural to try to identify a related, better mastered, problem or domain from which, hopefully, some useful information can be called upon for help. The emerging area of transfer learning is concerned with finding methods to transfer useful knowledge from a known source domain to a less known target domain. Our motivations are rooted in the recent advances in deep and metric learning where neural networks achieve very good results in image classification, speech recognition, and natural language processing, and even in difficult reasoning tasks such as *target sense verification* (TSV) [46], *case-base reasoning* [9, 13] and *knowledge graph bootstrapping* [3].

For supporting this line of research, we developed the platform ANNa¹ as a tool for showcasing the various applications of the analogy based framework, and making the corresponding models freely available to domain experts interested in applying AR. This initial platform is being extended in the framework of the ANR project AT2TA² to integrate different notions of analogy, different neural models to detect and solve analogies, as well as to process multimodal data for a variety of downstream tasks.

3.4 Axis 4. Computational Complexity

In addition to the research concerns above, the computational resources needed to algorithmically tackle the issues and tasks in the 3 previous axes constitute a continuous concern in the Orpailleur.

We propose general algebraic tools [32], that unify the usual extension of common computational problems under a common *semiring* formalism. Among the extensions concerned, one can mention the counting extensions (counting the number of solutions), the list extensions (allowing/forbidding certain solutions), and the cost extensions (that aims for a solution of minimal cost). Regarding the problems that benefit from this approach, our approach is very general, as it applies *any* problem, as long as we can define a “set of solution” for every instance: every problem of the class NP is therefore concerned. We note that this extensions enable to generalize significantly the scope of the problem treated without impact on the time complexity. Moreover, we build a new family of semirings, whose computational application is to enable to count the number of solutions of minimal weight. Despite how general this construction is, we apply it successfully to the design of FPT (fixed-parameter tractable) algorithms solving complex and undoubtedly practical problems, such as CONNECTED-DOMINATING-SET parameterized by the famous parameter *clique-width*, and the very general CONSTRAINT SATISFACTION PROBLEM (CSP) parameterized by the well-studied *tree-width*.

We illustrated the advantages of this framework by instantiating our general algorithmic approach on several classes of problems (e.g., the H -coloring problem and its variants), and showing that it improves the best complexity upper bounds in the literature for several well-known problems. In fact, while studying [30] the complexity of H -Coloring with respect to the parameters clique-width and the more recent component twin-width, we obtained two surprising linear bounds between these parameters, thus improving the previously known exponential and double exponential bounds. We also noticed that the technique used in order to obtain these bound can easily be adapted in order to derive similarly tighter bounds between the *linear clique-width* and the *total twin-width*. The former linear bounds entail natural approximations of component twin-width, by making use of the results known for clique-width. On the algorithmic side we targeted the richer problem of counting the number of homomorphism to H ($\#H$ -Coloring). The first algorithm uses a contraction sequence of the input graph G parameterized by the component twin-width of G . This leads to a positive FPT result for the counting version. The second uses a contraction sequence of the template graph H and here we instead measure the complexity with respect to the number of vertices in the input graph. Using our linear bounds we show that the latter always beats the previously best algorithm (based on clique-width) and we thus obtain the fastest general purpose algorithm for graph coloring.

Another line of research concerns discrete structures and the complexity of problems, in particular, graphs, polyhedra and combinatorics. Given any graph G , one can build a polytope $A(G)$, the graph associahedron of G that encodes the combinatorics of certain objects related to G . When the graph G is a path on n vertices, this polytope is the ubiquitous associahedron, whose vertices can be represented indifferently by the triangulations of a convex polygon with $n + 2$ vertices or the binary trees with n nodes. With different families of graphs, we retrieve other well-known polytopes: the associahedron of a complete graph is the permutohedron, the associahedron of a cycle is the cyclohedron, and the associahedron of a star is the stellohedron. In general, when G is an arbitrary graph, the vertices of $A(G)$ can be shown to correspond to the search trees on G (that are also sometimes called elimination trees), and the edges of $A(G)$ to the rotations between search trees.

The graphs consisting of the vertices and edges of a graph associahedron have been attracting a lot of attention due to their applications to data structures, sampling algorithms, and phylogenetics, among other subjects. The diameter and hamiltonicity properties of associahedra, cyclohedra, tree associahedra, chordal graph associahedra, caterpillar associahedra, and complete multipartite graph associahedra

¹A proof of concept is given in the morphological domain <https://anna.loria.fr/>

²Analogy: from Theory to Tools and Applications (AT2TA, 2023–2026), <https://at2ta.loria.fr/>.

have been considered. One of the most interesting computational complexity problem on this area is the following: Given a graph G and two search trees on G , what is the rotation distance between them? In a recent breakthrough paper (ICALP 2023), it is shown that the corresponding decision problem is NP-complete, by reduction from the balanced minimum (s, t) -cut in a graph. The question remains open in the case of the usual associahedron, or equivalently, when G is restricted to be the n -vertex path: It is not known whether the rotation distance between two binary search trees on n nodes can be computed in time polynomial in n , or whether the problem is NP-hard. In that case, the problem can also be cast as that of computing the flip distance between two triangulations of a convex polygon. It is known to be fixed-parameter tractable: For any constant k , we can decide in linear time if the flip distance is at most k . When G is a complete graph, hence when $A(G)$ is a permutohedron, the rotation distance between two search trees is simply the number of inversions between the two corresponding permutations, a quantity that can be computed using classical sorting algorithms.

4 Application domains

The methodologies developed by Orpailleur have several applications with direct societal, scientific and industrial impacts. Each application will typically adapt (some of) the KDD and ML/AI methodologies presented above in concrete application domains. Here we will illustrate some of these applications on three critical societal and industrial domains, namely, healthcare, agronomy and material industry.

4.1 Healthcare

Electronic Health Records (EHRs) enable secondary use of hospital historical data, and in particular the design and conduct of statistical clinical studies [39]. But, EHRs also demonstrated that they provide enough information on patients to train supervised algorithms with useful applications in healthcare, such as patient prioritization and diagnosis guidance [43, 37]. EHRs are complex data as they combine structured data (e.g., diagnostic codes, drug prescriptions, laboratory results) and unstructured data (e.g., clinical texts, images) in two dimensions: patient and time. These temporal and multidimensional aspects motivated with success the development of supervised approaches to learn patient representations from EHRs [40, 44]. This is similar to what is done in NLP with language representations, which embed language complex features e.g., word contexts, order, sentence syntax. For instance, in this healthcare context, sequences of words are replaced by sequences of patient-related events. In collaboration with Inria Paris and the IHU Imagine, we are currently investigating healthcare applications of the analogy based framework to patient representations learned from EHRs.

4.2 Agronomy

In collaboration with INRAE, CIRAD, LEPSE and the University of Caen-Normandie, we aim to better understand and predict the effects of isolated and recurrent heat waves on grain production considering both quantity and quality. Currently, we are focusing on three different crops nationwide, namely, sorghum, oilseed rape and wheat. The three partners identified heat stress sequences (intensity, duration, and position during reproductive phases) that impact plant production either affecting grain number, grain size or both, with or without affecting grain biochemical composition. By comparing effects of isolated heat stress with those of recurrent heat stress sequences, we observed that the plant final response to a sequence of recurring stressing events does not correspond to the sum of the individual responses to each event. Depending on the heat stress scenario, plant performance may be affected by recurrent heat stress otherwise than expected. Several hypotheses relying on published work mainly performed at the physiological and/or molecular scales may explain such patterns.

Orpailleur will apply ML techniques on the heterogeneous data previously acquired and analyzed to develop prediction modules at different temporal and spatial scales. Techniques that can consider the actual experience of the plants, such as long short-term memories (neural network) will be used to produce consistent prediction of yield and its components. These modules will be then used as (a) pure predictors, (b) knowledge providers for developing process-based formalisms to improve existing crop models, and (c) sub-components of the existing crop models. As far as possible we will try to take

advantage of the multi-species approach to develop common formalisms for the prediction of grain production in the three crops species.

4.3 Material science

Developing new materials remains a main challenge to follow the fast evolution of our society. Elaboration techniques must offer the possibility of developing such novel high-performance metals and alloys respecting environmental constraints. However, a mechanism-based tailoring of the performances requires constant improvements of experimental and theoretical techniques to unravel fundamental mechanisms controlling the macroscopic behavior. Plasticity is an important phenomenon, which is considered here. It is closely related to the mechanical strength and formability, and leads to progressive damage of components through non-reversible shape changes. Most of used materials are generally polycrystalline, where grains are separated from their neighbors by Grain Boundaries (GBs). Grains cannot be deformed independently because the cohesion between them must be maintained. Dislocations glide through grains and interact with each other or with the GBs acting as sinks, traps, obstacles, and sources of dislocations. GBs are very important from the mechanical properties point of view. Nowadays, we almost know how a dislocation interacts with a specific GB. But, understanding the collective response of several real GBs (contained in a real polycrystalline specimen) after receiving numerous dislocations is still a major scientific challenge. The difficulty becomes highly challenging when we consider the influence of the distribution of GBs, other types of interfaces, shape and orientation of grains, i.e. the microstructure.

In this context, our objective is to apply machine learning methods and explore a multi-level scale ranging from electron microscopy to micromechanics. At the micromechanics level, crystal plasticity models based on deep learning algorithms will be considered for suggesting specific microstructural parameters able to achieve targeted macroscopic mechanical properties. This project will have a major impact in current societal issues by enabling energy savings and limited costs associated with the tuning of microstructures targeting specific mechanical performances. We have developed a method which aligns two images (EBSD maps) produced by an electron microscope (SEM). The first (initial) image shows the surface of an unstressed material, whereas the second (final) image shows the deformation of the same surface when the material is loaded. The alignment method produces a deformation field which accounts for the displacement of each point of the material between the initial image and the final image. Acquiring images with a Scanning Electron Microscope takes a lot of time. As a result, the volume of the data available to us is very limited (a dozen images). This fact prevents us from relying on machine learning approaches for the time being. Consequently, the method developed attempts to find the optimal alignment between both images by optimizing a parametric geometric transformation. The initial results are satisfactory. Avenues for future improvement include: more intelligent initialization of the method (use of SIFT descriptors for instance), adaptation of the cost function to make it independent of certain measurement problems, production of simulated images for use in machine learning methods, devising methods that take into account knowledge of physical laws.

5 Social and environmental responsibility

In Orpailleur we are well aware that we develop methodologies that have immediate societal and environmental impacts, and that require ethical considerations. For instance, we have developed approaches to mitigate unintended biases in classifiers [34, 31] used in job application or in student admission processes, as well as to reduce stereotypical biases in language models [29]. Despite fulfilling their intended meaning, these results should be taken with a grain of salt when it comes to deploying such as technique in real-world scenarios.

Firstly, we define bias according to the metrics proposed in the literature, e.g., [5] and [41], but users of our proposed methodologies should make sure that these bias metrics are compatible with their requirements and ethical expectations. Moreover, the bias measures used only reflect some indicators of undesirable biases or stereotypes, and users should avoid using our systems as proof or as a guarantee that their models are unbiased without extensive study [38, 35]. This is particularly relevant in the case of FixOut that is currently under startup maturation.

Secondly, users and practitioners should be aware of the considerable financial and carbon footprints. This is particularly relevant for our approaches to debias large language models (LLMs), which require both training and experimenting with LLMs [33], and in Orpailleur we recall members of the environmental footprint, thus advocating limitations to the massive usage of such models to reasonable amounts.

It is also noteworthy that Lydia Boudjeloud is a member of the “Commission de parité” du LORIA. The commission members are raising awareness throughout the unit, not only among doctoral students and undergraduates, but also among the youngest members of the community, by promoting initiatives aimed at primary schools and secondary schools, and by proposing courses of action to promote parity.

6 New softwares, platforms, open data

6.1 New software

6.1.1 FixOut

- **Name:** FixOut: algorithmic inspection for trustworthy AI
- **Keywords:** algorithmic fairness, fairness assessment, fairness mitigation, explanations in artificial intelligence, XAI, feature importance, feature dropout, ensemble classifier.

- **Functional Description:**

FixOut is a Python package that checks, explains and mitigates algorithmic unfairness in machine learning models (particularly binary classifiers). It is built on top of the original FixOut framework **??**. The package has two main components: (1) fairness assessment and (2) unfairness mitigation.

1. Given a pre-trained model and a training dataset, the first component is able to (a) produce a ranking of features based on their importance to the model’s outputs; (b) calculate standard fairness metrics (group fairness); (c) evaluate the reliance between sensitive features and non-sensitive features; (d) compare the pre-trained model with previous models w.r.t. fairness metrics.
2. If the pre-trained model is deemed unfair, FixOut triggers the component responsible for the unfairness mitigation. The original FixOut framework is then employed to produce a fairer model. It thus applies a feature dropout approach that produces a pool of simplified classifiers that are then aggregated into an ensemble classifier.

Finally, FixOut produces a comprehensive report that consolidates all the findings obtained at each component (1) and (2) and then displays them in a web interface, so that all stakeholders can be aware of the current state of the unfairness assessment and mitigation.

- **Partnership:** SATT Sayens, Incubateur Lorrain
- **URL:** <http://fixout.tech>

6.1.2 ANNa

- **Name:** ANNa - Analogy and Neural Network Application
- **Keywords:** Analogy, neural network, morphological analogy, classification model, retrieval model, generation model.
- **Functional Description:** The ANNa platform serves as a comprehensive showcase and introduction to analogies, with a specific focus on morphological analogies. The primary purpose of this platform is to demonstrate the outcomes of our work in the realm of analogies, particularly emphasizing morphological aspects. We have trained analogy models for more than 11 different languages, employing three key models within the platform: the Classification Model, Retrieval Model, and Generation Model.

- **Classification Model (ANNc):** ANNc, or Analogy Neural Network Classification, is designed to classify given analogies as either valid or not, providing a confidence score for the model's prediction. Input to this model comprises four words arranged in an analogy format (e.g., A:B::C:D), and the model determines the validity of the analogy along with a confidence measure.
 - **Retrieval Model (ANNr):** ANNr, or Analogy Neural Network Retrieval, is responsible for retrieving the closest word from the embedding space. For instance, when presented with an incomplete analogy (A:B::C:?), where the fourth term (D) is unknown, the model suggests the closest word, completing the analogy as (A:B::C:D).
 - **Generation Model (ANNa for Generation):** ANNa's Generation Model is dedicated to generating the missing term (D) based on the given analogy (A:B::C:?). After computation, the model produces the generated result, completing the analogy as (A:B::C:D).
- **Integration and Microservices:** All three models, ANNc, ANNr, and ANNa for Generation, have been seamlessly integrated into the ANNa platform. The integration involves the deployment of numerous microservices operating from the platform's backend, ensuring the smooth execution and interaction of these models.
 - **URL:** <https://anna.loria.fr/>

6.1.3 Caspailleur and Paspailleur

- **Name:** Caspailleur and Paspailleur
- **Keywords:** Formal Concept Analysis, concepts, closed sets, implications, pattern structures, data analysis.
- **Functional Description: Caspailleur** Caspailleur is a Python package for mining many concise data representations (<https://pypi.org/project/caspailleur/>). This software is developed in the framework of the SmartFCA ANR Research Project.
- **Functional Description: Paspailleur** Paspailleur is a companion project whose objective is to implement an extension of the Caspailleur functionalities to take into account complex data such as numbers, strings, trees, and graphs (<https://pypi.org/project/paspailleur/>).
For knowing more about Caspailleur and Paspailleur, see <https://github.com/EgorDudyrev/espailleur/blob/main/>.
- **Partnership:** see SmartFCA ANR Project.
- **URL:** <https://pypi.org/project/caspailleur/> and <https://pypi.org/project/paspailleur/>.

6.2 New platforms

6.2.1 AT2TA-Nexus

The **AT2TA-Nexus** platform consolidates developed and upcoming tools, functioning as a showcase for the innovative solutions of the ANR AT2TA project. Our ongoing project focuses on creating a cohesive platform tailored to this objective.

- **Key Components:** The **AT2TA-Nexus** platform will seamlessly integrate ANNa, a significant tool for analogies and neural network applications. Other tools currently in development will also be incorporated.
- **Objective and Functionality:** The **AT2TA-Nexus** website, a dynamic web application, aims to provide users with an easy and intuitive experience. As a comprehensive showcase, it offers insights into diverse projects consolidated within the unified platform.

- **Current and Future Development:** Under active development, the **AT2TA-Nexus** platform is evolving to include various tools from the ANR/AT2TA project. The integration of ANNa is underway, and plans for additional tools emphasize our commitment to creating a user-friendly space for innovative applications.

6.3 Open data

- L. Jarnac, P. Monnin, M. Couceiro. Wikidata Thematic Subgraph Selection, 2023. <https://zenodo.org/records/8091584>
- E. Marquer, M. Couceiro, S. Alsaidi, A. Decker. Siganalogies - morphological analogies from Sigmorphon 2016 and 2019. [doi:10.12763/MLCFIE](https://doi.org/10.12763/MLCFIE)

7 New results

7.1 Axis 1. Exploratory and hybrid knowledge discovery

Participants: Brieuc Conan-Guez, Miguel Couceiro, Alain Gély, Alexandre Blansch , Lydia Boudjeloud-Assala, Amedeo Napoli, Fr d ric Pennerath, Yannick Toussaint.

7.1.1 Connexions between FCA and Decision Trees (description quivers)

We have introduced and studied special structures named “description quivers” as compact representations of concept lattices and respective ensembles of decision trees. Formally, description quivers are directed multigraphs where vertices represent concept intents and (multiple) edges represent generators of intents. We have studied some properties of description quivers and shed light on their use for describing state-of-the-art symbolic machine learning models based on decision trees. We also argue that a concept lattice can be considered as a cornerstone in constructing an efficient machine learning model. We show that the proposed description quivers allow us to fuse decision trees just as we can sum linear regressions, while proposing a way to select the most important rules in decision models, just as we can select the most important coefficients in regressions.

7.1.2 A study of the closure algorithms and the discovery of implication bases

Implications in Formal Concept Analysis (FCA), Horn clauses in Logic, and Functional Dependencies (FDs) in the Relational Database Model, are very important dependency types in their respective fields. Moreover, they have been proved to be equivalent from a syntactical point of view. Then notions and algorithms related to one dependency type in a field can be reused and applied to another dependency type in the other field. One of these notions is that of “cover”, also known as a “basis”, i.e., a compact representation of a complete set of implications, of functional dependencies, and Horn clauses. Although the notion of cover exists in the three fields, the characterization and the related uses of a cover are different. We study and compare, from an FCA perspective, the principles on which rely the most important covers in each field. We also intend to study some open questions that are of interest in the three fields, and especially to the FCA community, in particular about the design and the use of such implication bases.

7.1.3 A study of data complexity from an FCA point of view

We are interested in data complexity and how it can be measured. Thus we introduced different indices for measuring the complexity of a dataset in terms of Formal Concept Analysis (FCA). We extend the lines of the research about the “closure structure” and the “closure index” based on minimum generators of intents (aka closed itemsets). We would try to capture statistical properties of a dataset, not just extremal characteristics, such as the size of a passkey. For doing so we introduce an alternative approach

where we measure the complexity of a dataset w.r.t. five significant elements that can be computed in a concept lattice, namely intents (closed sets of attributes), pseudo-intents, proper premises, keys (minimal generators), and passkeys (minimum generators). Then we define several original indices allowing us to estimate the complexity of a dataset. Moreover we study the distribution of all these different elements and indices in various real-world and synthetic datasets. Finally, we investigate the relations existing between these significant elements and indices, and as well the relations with implications and association rules.

7.2 Axis 2. Explainability, transparency and algorithmic fairness

Participants: Briec Conan-Guez, Miguel Couceiro, Mohammed Fellaji, Amedeo Napoli, Frédéric Pennerath, Claire Theobalt.

7.2.1 Fairness Notions and Related Tensions

Automated decision systems are increasingly used to take consequential decisions in problems such as job hiring and loan granting with the hope of replacing subjective human decisions with objective machine learning (ML) algorithms. However, ML-based decision systems are prone to bias, which results in yet unfair decisions. Several notions of fairness have been defined in the literature to capture the different subtleties of this ethical and social concept (e.g., statistical parity, equal opportunity, etc.). Fairness requirements to be satisfied while learning models created several types of tensions among the different notions of fairness and other desirable properties such as privacy and classification accuracy.

We have surveyed [5] the commonly used fairness notions and discussed the tensions among them with privacy and accuracy, as well as reviewed different methods to address the fairness-accuracy trade-off (classified into four approaches, namely, pre-processing, in-processing, post-processing, and hybrid). This study is consolidated with experimental analysis carried out on fairness benchmark datasets to illustrate the relationship between fairness measures and accuracy in real-world scenarios.

7.2.2 Statistical approaches to detect disparity prone features in group fairness settings

The use of machine learning models in decision support systems with high societal impact raised concerns about unfair (disparate) results for different groups of people. When evaluating such unfair decisions, one generally relies on predefined groups that are determined by a set of features that are protected by law or considered to be sensitive. However, with respect to the latter, such an approach is subjective and does not guarantee that these features are the only ones to be considered as sensitive nor that they entail unfair (disparate) outcomes. We have proposed a preprocessing (statistical) approach to address the task of automatically recognizing disparity prone features that does not require a trained model to verify unfair results. Our proposal [2] is based on the Hilbert-Schmidt independence criterion, which measures the statistical dependence of variable distributions. We hypothesize that if the dependence between the label vector and a candidate is high for a sensitive feature, then the information provided by this feature will entail disparate performance measures between groups. Our empirical results attest our hypothesis and show that several features considered as sensitive in the literature do not necessarily entail disparate (unfair) results.

We have recently extended this basic framework in several ways. Precisely, we can now account for coalitions (combinations) of features, that may themselves entail disparate outcomes, and we combine them with their importance to the outcomes, which facilitates coalition interpretation and the measurement of their impact in the performance metrics (for instance, fairness metrics). Despite being a general framework, for the time being we only illustrated its feasibility by taking Shapley values. As future work it would be interesting to consider further combinations of features and their impact on fairness metrics. Indeed, we looked at disparities from a global perspective, but a fine grained analysis may reveal hidden disparities. Moreover, we are currently extending our framework to multiclass classification and regression settings.

7.2.3 Clarity

In the thesis [21] of Claire Théobald, defended in 2023, we tackled the problem of the explainability of deep neural networks. She developed a method called Clarity [45] to produce quality counterfactual explanations. A counterfactual explanation expresses the minimal realistic modifications that must be made to an input so that a classifier predicts a new target class. For instance, in the case of tabular data, if a classifier suggests refusing a loan to an individual, a counterfactual explanation is a change in that person's situation (for instance the discontinuation of certain subscriptions) which leads to acceptance of the loan. Claire Théobald worked on visual counterfactuals. Dealing with images instead of tabular data is particularly challenging. Indeed, a fundamental issue is that the method tries to generate a realistic image representing a high-level concept by modifying low-level data (pixels).

Clarity relies on two types of neural networks in order to overcome this problem: generative models (variational auto-encoders VAE) and Bayesian Neural Networks (BNN). VAE are used to encode images in a low-dimensional latent space. Clarity modifies the image in this latent space rather than in the high-dimensional pixel space. This approach ensures noise-free high level modifications to the original image. A special feature of Clarity is that the VAE encoder forms the first layers of the classifier. Therefore, Clarity's classifier can be seen as explainable by design. This architecture choice improves the algorithm convergence compared to more classical approaches (even those based on VAE). Computation times are also greatly reduced.

Quality counterfactual explanations must be realistic. In the case of the loan example, suggesting a 200% increase in income is for instance unrealistic. To quantify and improve realism, Clarity makes use of a Bayesian neural network, which form the last layers of the classifier. Bayesian models are able to provide uncertainty estimates associated to their predictions. These uncertainties can be directly interpreted as measures of realism: the model is uncertain for unrealistic examples (out of distribution examples). Thanks to VAEs and BNNs, Clarity is fast and produces quality explanations. This has been validated on various datasets. One of them is the dataset Celeba, which contains a large number of celebrity faces. For Celeba, the classification task consists in predict the color of the hair appearing in each image. In the event that the classifier indicates that a face has brown hair, producing a counterfactual explanation means modifying the image so that the classifier considers another color (blond, for example) to be more likely. On this dataset, Clarity delivers more realistic results than competing methods.

7.2.4 Bayesian Deep Learning and Reference Priors

In the framework of Mohammed Fellaji's thesis, we investigate the fundamental causes of calibration problems, and how to remedy them. More precisely, we are conducting a systematic study, where we are showing that even on elementary classification problems (MNIST, CIFAR10, etc.), Bayesian deep networks produce poorly calibrated epistemic uncertainty measures, which do not respect two expected principles, namely the decrease in uncertainty as the amount of training data increases or as the size of the model decreases. This work was published at the French machine learning conference (CAP). Afterwards, theoretical development has enabled us to link this calibration problem to the theory of reference priors, and to deduce a particular form of prior that largely resolves the observed shortcomings. This work is currently being evaluated.

7.3 Axis 3. Analogy based reasoning

Participants: Miguel Couceiro, Alain Gély, Lucas Jarnac, Esteban Marquer, Cherif-Hassan Noursadine.

7.3.1 Galois theory for analogical classifiers

Analogical proportions can be seen as 4-ary relations that read "A is to B as C is to D". Recent works have highlighted the fact that such relations can support a specific form of inference, called analogical inference. This inference mechanism was empirically proved to be efficient in several reasoning and classification tasks. In the latter case, it relies on the notion of analogy preservation. In [6] we explored

this relation between formal models of analogy and the corresponding classes of analogy preserving functions, and we establish a Galois theory of analogical classifiers. We illustrated the usefulness of this Galois framework over Boolean domains, and we explicitly determined the closed sets of analogical classifiers, i.e., classifiers that are compatible with the analogical inference, for each pair of Boolean analogies.

7.3.2 Case-base reasoning and analogical transfer.

We introduced an energy-based approach [13] to measuring the competence of a case base for machine learning tasks such as case prediction, classification, and analogical transfer. This competence approach differs from prior approaches proposed in the literature as it relies on the optimization of a global compatibility indicator between two similarity measures, one on the situation space and the other on the outcome space. We showed empirically that this notion of competence is tightly related to performance for a case-based classification task, in the sense that the competence of a source case is positively correlated to its ability to reduce the energy of correct outcomes and to increase the energy of incorrect outcomes. We analyzed both quantitatively and qualitatively the behavior of this competence-based approach on different datasets (with substantially different distributions) and taking into account different classification frontiers and loss functions. Moreover, we analyzed its robustness with respect to different reference and initial cases. These results suggest the strong potential of this energy-based framework for guiding case base maintenance, providing an alternative to existing methods. The empirical and thorough comparison between the former and the latter will constitute one of the topics to be investigated. We will also devote efforts to establish theoretical guarantees of these empirical results.

7.3.3 Knowledge Graph Bootstrapping via Zero-Shot Analogical Pruning

Knowledge Graph Construction (KGC) can be seen as an iterative process starting from a high quality nucleus that is refined by knowledge extraction approaches in a virtuous loop. Such a nucleus can be obtained from knowledge existing in an open KG like Wikidata. However, due to the size of such generic KGs, integrating them as a whole may entail irrelevant content and scalability issues.

We proposed [3] an analogy-based approach that starts from seed entities of interest in a generic KG, and keeps or prunes their neighboring entities. We evaluated our approach on Wikidata through two manually labeled datasets that contain either domain-homogeneous or -heterogeneous seed entities. We empirically showed that our analogy based approach outperforms LSTM, Random Forest, SVM, and MLP, with a drastically lower number of parameters. We also evaluated its generalization potential in a transfer learning setting. These results advocate for the further integration of analogy-based inference in tasks related to the KG lifecycle.

7.4 Axis 4. Computational complexity

Participants: Miguel Couceiro, Ambroise Baril, Mario Valencia.

7.4.1 Fine-grained and parameterized complexity of graph coloring problems.

In [30], we showed that component twin-width and clique-width are linearly equivalent, and used this equivalence to provide an FPT algorithm for H -COLORING parameterized by component twin-width. We also proposed two algorithms for solving $\#H$ -COLORING. The first uses a given optimal contraction sequence of the input graph G to solve $\#H$ -COLORING in FPT time parameterized by component twin-width. The second uses a contraction sequence of the template graph H and beats the clique-width approach for solving $\#H$ -COLORING (with respect to the size of G).

This work still raises some unsettled questions, *e.g.*, (i) Are the bounds relating component twin-width and clique-width that we provided tight in general?³ (ii) Are the running times of our FPT algorithms

³We showed that they are tight for the class of cographs with at least one edge.

with respect to certain parameters optimal?⁴ (iii) Can we extend our results concerning solutions of $\#H$ -COLORING to arbitrary binary constraints (BINARY-CSPs)?⁵

7.4.2 Rotation distance of split graph associahedra is polynomial

Given a graph G , we can construct a graph $A(G)$ called the graph associahedra of G , where the vertices of $A(G)$ can be shown to correspond to the search trees on G (that are also sometimes called elimination trees), and the edges of $A(G)$ to the rotations between search trees on G . A major problem on graph associahedra is the computational complexity of the distance function, that is, the minimum number of rotations needed to transform any elimination tree into another one. When G is a complete graph, hence when $A(G)$ is a permutahedron, the rotation distance between two search trees is simply the number of inversions between the two corresponding permutations, a quantity that can be computed using classical sorting algorithms. The permutahedron was the only graph associahedra family for which the rotation distance problem was known to be polynomial. We show recently (Eur. J. of Combinatorics, Vol. 118, May 2024) that the rotation distance between two search trees on a complete split graph with n vertices can be computed in time $O(n^{2+o(1)})$. A complete split graph G is obtained from the complete graph by selecting a non-empty proper subset Q of its vertices and by removing all the edges between two vertices from Q . If Q is a singleton, then G is complete and $A(G)$ is the permutahedron. On the other hand, if Q contains all the vertices of G but one, then G is a star and $A(G)$ is the stellohedron. For arbitrary graphs, a recently result (ICALP 2023) shows that the rotation distance of graph associahedra is an NP-complete problem.

7.4.3 Exact distance Kneser graphs

Let G be a connected graph. Given two vertices a, b in G , $dist(a, b)$, the *distance* between a and b , is defined as the length of the shortest path in G joining a to b . The diameter of G is defined as the maximum distance between any pair of vertices in G . The concept of exact distance- d graph, where d is a positive integer, has been introduced by Nešetřil and Ossona de Mendez. Formally, if $G = (V, E)$ is a graph, then the exact distance- d graph $G_{=d}$ of G is the graph with vertex set V and where two vertices a, b in $G_{=d}$ are adjacent if and only if they are at distance exactly d in G . Note that $G_{=1} = G$.

The main focus in early investigations of exact distance graphs was on their chromatic number. The exact distance- d graphs have been much earlier considered for hypercube graphs in the frame of the so-called cube-like graphs. More recently, Brešar et al. considered the structure of exact distance- d graphs of some graph products. They remarked that the concept of exact distance graphs is not only interesting because of the chromatic number, but also as a general metric graph theory concept.

Let k, r be positive integers and let $[2k+r] = \{1, 2, \dots, 2k+r\}$. Let $[2k+r]^k$ be the set of k -subsets of $[2k+r]$. The Kneser graph $K(2k+r, k)$ is the graph with vertex set $[2k+r]^k$ and where two k -subsets $A, B \in [2k+r]^k$ are joined by an edge if $A \cap B = \emptyset$. Note that $K(5, 2)$ is the well-known Petersen graph. It is easy to show that the Kneser graph $K(2k+r, k)$ is a connected regular graph having $\binom{2k+r}{k}$ vertices of degree $\binom{k+r}{k}$. There exist some generalizations to Kneser-like graphs in the literature. Let $K(n, k, s)$ be the generalized Kneser graph, i.e. the graph with vertex set $[n]^k$ and the edges connecting all pairs of vertices with intersection smaller than s . Denote by $J(n, k, s)$ the generalized Johnson graph, i.e. the graph with the same vertex set $[n]^k$ and edges connecting all pairs of vertices with intersections exactly s . These graphs are quite popular objects in combinatorics. The chromatic number of the generalized Kneser graph was studied by Frankl and Füredi for fixed k and s . Diameters of $K(n, k, s)$ and $J(n, k, s)$ are known.

We have studied the *exact distance- d* Kneser graphs $K_{=d}(2k+r, k)$, with k, r positive integers. Let D denote the diameter of the Kneser graph $K(2k+r, k)$ and let d be a fixed integer with $1 \leq d \leq D$. We have obtained a complete characterization of the adjacency relation on $K_{=d}(2k+r, k)$ in terms of the cardinality of the intersection of its vertices. In fact, we prove that two vertices $A, B \in K_{=d}(2k+r, k)$ are adjacent if and only if their cardinality intersection lies between an integer interval depending on d, k , and r . We have also computed the distance of any two vertices in $K_{=d}(2k+r, k)$ and its diameter. All these results will be submitted to a journal of combinatorics in few days.

⁴Ganian *et al.* [36] showed that algorithms relying on clique-width to solve H -COLORING are optimal, and we have a similar optimality result with respect to treewidth.

⁵The notion of component twin-width does indeed generalize naturally to both instances and templates of a BINARY-CSP.

8 Bilateral contracts and grants with industry

8.1 Bilateral contracts with industry

8.1.1 Happiso

- Company: Happiso
- Participants: Miguel Couceiro, Morgane Ruiz-Huidobro
- Duration: 3 years
- Cifre: Exploiting explanations and fairness of neural approaches for a decision support system

8.1.2 Orange

- Company: Orange
- Participants: Miguel Couceiro, Lucas Jarnac
- Duration: 3 years
- Cifre: Réconcilier des connaissances incertaines pour améliorer un graphe de connaissances

8.2 Grants with industry

8.2.1 Kesaio (2022–2024)

The objective of the KESAIO project is to propose a semi-automated process for building and expanding a domain ontology as a component of a knowledge-based system. The input is a dataset of linked data to be mined to build a domain ontology. The output should be a part of a domain ontology to be used by specialists in possibly critic situations for guiding and helping decision making. In this project, we are making use of methods related to knowledge discovery, knowledge representation, and reasoning.

The Kesaio project is carried out with the startup Airudit in Bordeaux (see <https://www.airudit.com/fr-FR/home>).

8.3 Transfer with Industry

8.3.1 FixOut start-up

The FixOut start-up provides software as a service (SaaS) to help organisations audit their ML models. FixOut's main tool is a software centred on algorithmic fairness and explainability. It is designed to make it easier the assessment and reduction of unintended unfair algorithmic outcomes, so organisations can get prepared for the upcoming new European AI legislation. In order words, to check whether ML-based systems are compliant with the AI Act by

- helping organisations to assess and reduce unintended bias without reducing performance, and improve the quality of their ML models;
- protecting organisations (and unprivileged sub-populations) from the risks associated with unintended biases in their ML-based systems.

FixOut's technology is the result of the research on fairness and explainability carried out by the Orpailleur team since 2020. Before becoming a start-up project, the research behind FixOut resulted in published papers [31, 5] and a doctoral thesis. Currently in its maturation phase (since 2023) with the support of SATT Sayens, the first version of FixOut's main tool was registered⁶ in the *Agence pour la Protection des Programmes* (APP). The brand FixOut is also registered under the number FR 4904168.

Loria contact: Miguel Couceiro (scientific consultant)

Contact: Guilherme Alves

⁶The APP deposit certificate for "FixOut (Tabular version) Version 1.0 of 31.01.2022" dated of 28.02.2022 bears the reference IDDN.FR.001.090003.000.S.C.2022.000.20900

9 Partnerships and cooperations

9.1 International research visitors

9.1.1 Visits of international scientists

- Yves Lepage (Professor), University of Waseda, Japan. Sabbatical: Sep. 2023 - April 2024.
- Marc Munar (Dr.), Soft Computing, Image Processing and Aggregation (SCOPIA) research group, Dept. of Mathematics and Computer Science, University of the Balearic Islands, Palma, 07122, Spain. Research stay: Sep.-Dec. 2023.
- Agustina Ledezma (Dr.), Instituto de Matemática Aplicada San Luis, Universidad Nacional de San Luis and CONICET, San Luis, Argentina. Research stay: May-June 2023.
- Ana Gargantini (Dr.), Facultad de Ciencias Exactas y Naturales Universidad Nacional de Cuyo, Mendoza, Argentina. Research stay: October 2023.
- Daniel Quiroz (Assistant Professor), Valparaiso university, Chile. Research stay: June 2023.

9.1.2 Visits to international teams

Miguel Couceiro:

- **Visited institution:** Instituto Superior Técnico, Universidade de Lisboa
Country: Portugal
Dates: March and July
Context of the visit: Research presentations and discussions at the Dep. Mathematics and the laboratory INESC-ID for project preparation.
Mobility program/type of mobility: research stay
- **Visited institution:** Charles University of Prague
Country: Czechia
Dates: September
Context of the visit: Preparation of the new Erasmus Mundus ENLP
Mobility program/type of mobility: research stay
- **Visited institution:** University of Malta
Country: Malta
Dates: June
Context of the visit: Annual Meeting of the Erasmus Mundus LCT
Mobility program/type of mobility: short stay

9.2 European initiatives

TAILOR

Title: Foundations of Trustworthy AI - Integrating Reasoning, Learning and Optimization

Duration: 2020-2024

Coordinator: Fredrik Heintz (Linköpings Universitet)

Partners: 53 institutions and companies all across Europe

Loria contact: Emmanuel Vincent

Summary: TAILOR aims to bring European research groups together in a single scientific network on the Foundations of Trustworthy AI. The four main instruments are a strategic roadmap, a basic research programme to address grand challenges, a connectivity fund for active dissemination, and network collaboration activities. Miguel Couceiro is involved in fairness, equity, and justice by design research in WP3. He is also member of the scientific board of the *TAILOR Connectivity fund*

Project website: [TAILOR](#)

9.3 National initiatives

9.3.1 ANR

ANR InExtenso

Title: Intrinsic and Extrinsic evaluation of biases in large language models

Duration: 2023-2027

Coordinator: Karen Fort

Partners:

- Laboratoire lorrain de recherche en informatique et ses applications (LORIA)
 - Laboratoire Interdisciplinaire des Sciences du Numérique (LISN)
Local coordinator: Aurélie Névél
 - CHU de Rouen
Local coordinator: Stéfan Darmoni

Loria contact: Karen Fort

Summary: Large Language Models (LLM) are the Swiss army knife of today's Natural Language Processing (NLP). They often outperform the state-of-the-art on benchmarks commonly used in the field for tasks such as part-of-speech tagging, text classification and named-entity recognition, thus paving the way to a myriad of end-user applications. However, it has been shown that LLM exhibit major ethical issues including significant environmental impact, mirroring and amplification of stereotyped biases, which in turn have a disproportionate impact on historically disadvantaged social groups. It is urgent to address the social impact of NLP as the applications we develop, such as chatGPT, are now directly made available to end-users. The detection and mitigation of biases has therefore become an active area of research in the past few years, focusing mainly on Masked Language Models (MLM) such as BERT in English and the north American social context. Several sources of bias were identified in the NLP pipeline, however the inter-connection between sources and overall impact of each source on downstream applications remains unclear. In this project, we want to observe the entire pipeline, from the intrinsic point of view (within the model itself), to the pre-training task point of view (in the case of auto-regressive LLM, text generation), on to some real-world downstream applications. We chose to focus on two types of medical applications: mental illness diagnosis help and information extraction from clinical records for public health purposes such as patient enrollment into clinical trials. The project will provide corpora and methods for a global evaluation of bias in LLM in French as well as studies to further the understanding of biases in clinical NLP pipelines and the environmental impact of the integration of these models in digital health. Miguel Couceiro contributes as the co-leader of WP2 on intrinsic evaluation and debiasing.

Project website: [InExtenso](#)

ANR AT2TA

Title: Analogies: From Theory to Tools and Applications

Duration: 2023-2026

Coordinator: Miguel Couceiro (LORIA)

Partners: • Laboratoire lorrain de recherche en informatique et ses applications (LORIA)

- Université Paul Sabatier, CNRS, IRIT
Local coordinator: Stergos Afantenos
- Inria Paris
Local coordinator: Adrien Coulet
- IHU Imagine
Local coordinator: Nicolas Garcelon
- Infologic
Local coordinator: Mehdi Kaytoue
- Université Côte d'Azur, Inria, CNRS, I3S
Local coordinator: Pierre Monnin

Loria contact: Miguel Couceiro

Summary: Analogical reasoning is a remarkable capability of human reasoning. Analogical proportions are statements of the form “A is to B as C is to D”. They are the basis of analogical inference that has been used in machine learning (ML) tasks such as classification, decision making, and automatic translation with competitive results. Analogical extrapolation can solve hard reasoning tasks, such as IQ tests, and support data augmentation when learning models with few labeled samples. What makes analogical inference special is its unique ability to simultaneously process similarities and dissimilarities. Analogical reasoning links the two main axes of AI (knowledge representation and reasoning, and machine learning), and contributes to the transparency and explainability of AI as it is close to human reasoning and enables explanations based on examples and counter-examples. This motivates our efforts to develop an analogy-based ML framework and to demonstrate its usefulness in real world applications. We will explore analogical reasoning for transfer learning and case-based reasoning, where the idea is to take advantage of what has been learned on a source domain in order to improve the learning process in a target domain related to the source domain. Suitable representations are the key to transfer the analogy-based framework to other settings and to handle different object types. This asks for a thorough study of representation spaces for analogy-based frameworks with different object types, not only textual and tabular, but also complex and structured, e.g., patient data, knowledge graphs and abstract syntax trees. As the final goal, the AT2TA project aims to provide an open access platform to detect, solve, and reason with analogies, illustrated by noteworthy applications in NLP, medical sciences, as well as in knowledge management and software engineering, which have a major impact in industry.

Project website: [AT2TA](#)

ANR MAMIENOVA

Title: Machine leArning for MIcromEchanics: a NOVel Approach

Duration: 2023-2026

Coordinator: Antoine Guitton (LEM3)

Partners: • Laboratoire d'Etude des Microstructures et de Mécanique des Matériaux (LEM3)

- Université de Lille
Local coordinator: Mussi Alexandre
- Institut national de la recherche en informatique et automatique (LORIA)
Local coordinator: Lydia Boudjeloud-Assala

Loria contact: Lydia Boudjeloud-Assala

Summary: The ultimate objective of materials science is to be able to adapt microstructures to reach desired properties. However, no consistent constitutive models were made to date essentially because of the need to statistically link the microscopic and macroscopic scales. In this project, we propose an original methodology where a crystalline plasticity code will be coupled to a supervised learning algorithm to obtain a system capable of suggesting the distribution of operating mechanisms in a polycrystal with its microstructural parameters in order to obtain desired macroscopic mechanical properties. This new model resulting from the learning process will be instructed from a large set of experimental data obtained by scanning electron microscopy and translated into an input-output system. This project will have a major impact in current societal issues by enabling energy savings and limited costs associated with the tuning of microstructures targeting specific mechanical performances.

ANR RICOCHETS

Title: Resilience to recurrent Heat Stresses in plants

Duration: 2023-2026

Coordinator: Christine Granier (AGAP)

Partners:

- Institut national de la recherche en agronomie (AGAP)
- Université de Caen Normandie (EVA)
Local coordinator: Sophie Brunel-Muguet
- Institut national de la recherche en informatique et automatique (LORIA)
Local coordinator: Miguel Couceiro
- Laboratoire d'Écophysiologie des Plantes sous Stress environnementaux (LEPSE)
Local coordinator: Boris Parent
- Biologie du Fruit et Pathologie (BFP)
Local coordinator: Yves Gibon

Loria contact: Miguel Couceiro

Summary: Crop production data combined with future climate projections can be used to estimate how production, including crop yield but also quality, will be affected by global changes. However, these trends are based on relationships between average environmental conditions and average plant responses. Because there is increasing evidence that the magnitude and frequency of extreme weather events, such as heat waves, are increasing, there is a need to (1) better understand how recurrent environmental stresses within a season affect plant production and (2) incorporate plant responses to such stresses in crop models. RICOCHETS is based on the strong assumption that plant response to recurrent heat stresses, i.e., the succession of heat stresses events separated by non-stressing ones, may not match the addition of individual responses to each event. According to several studies mainly performed at the physiological and/or molecular levels, this could be explained by several reasons. First, a certain form of plant recovery, which can lead to compensation or even over-compensation, following the stressing period. Second, plants subjected to a first stress can be more resilient or less resilient to a second period of stress because they have been respectively prepared ('primed') or too much affected by the first stressing event. Combining a data-driven (knowledge discovery in database) and a concept-driven approach (multi-scale analyses of plant responses during recurrent heat stress sequences imposed at specific stages of plant development), RICOCHETS aims to better understand, model and predict the effects of recurrent heat stresses on plant growth, development and production.

ANR SmartFCA

Title: Formal Concept Analysis: A Smart Tool for Analyzing Complex Data

Duration: 2022-2025

Coordinator: Marianne Huchard (LIRMM)

Partners:

- Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM)
- Laboratoire lorrain de recherche en informatique et ses applications (LORIA)
Local coordinator: Amedeo Napoli
- Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA)
Local coordinator: Sébastien Ferré
- Agroécologie et Intensification Durable des cultures Annuelles (AIDA)
Local coordinator: Pierre Martin
- Laboratoire des sciences de l'Ingénieur, de l'Informatique et de l'Imagerie (ICube)
Local coordinator: Florence Le Ber
- Laboratoire Informatique Image et Interaction (L3i)
Local coordinator: Karell Bertet
- Infologic
Local coordinator: Mehdi Kaytoue

Loria contact: Amedeo Napoli

Summary: 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.

Project website: [SmartFCA](#)

ANR AstroDeep

Title: Analysis of large astronomical datasets with machine learning

Duration: 2019-2023

Coordinator: Eric Aubourg (APC)

Partners:

- Astroparticule et Cosmologie (APC)
- Laboratoire lorrain de recherche en informatique et ses applications (LORIA)
Local coordinator: Amedeo Napoli
- DAp Département d'Astrophysique/Institut de recherche sur les lois fondamentales de l'Univers
Local coordinator:

Loria contact: Amedeo Napoli

Summary: Astronomical surveys planned for the coming years will produce data that present analysis challenges not only because of their scale (hundreds of petabytes), but also by the complexity of the measurement challenges on very deep images (for instance subpercent-level measurement of colors or shapes on blended objects). Machine learning techniques appear very promising: once trained, they are very fast, and excel at extracting features from images. Preliminary results

using modified variational auto-encoders as deblenders are very encouraging. We aim to develop machine learning techniques that can be applied directly on multi-bandpass, multi-instrument individual images to address the key observation challenges without going through the traditional steps of image stacking, explicit deblending, catalog generation, that lose information at each stage. We will target as main objectives the measurement of weak gravitational lensing (through a local average shape of objects), and of photometric redshifts (producing a bayesian $p(z)$ probability for each object of a blend). To achieve these goals, we will develop new deblending techniques, and will use and improve bayesian deep learning techniques in order to achieve a proper and consistent handling of uncertainties. Those techniques will help leverage the observation capabilities of future surveys like LSST, Euclid and WFIRST, and will allow a joint analysis of their data.

Project website: [AstroDeep](#)

10 Dissemination

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

General chair, scientific chair

- Miguel Couceiro co-chaired the IJCAI Workshop *Interactions between Analogical Reasoning and Machine Learning* (IARML 2023), China.
- Miguel Couceiro and Esteban Marquer co-chaired the ICCBR Workshop *Analogies: from Theory to Applications* (ATA 2023), Scotland.
- Amedeo Napoli is a co-organizer with Sergei O. Kuznetsov and Sebastian Rudolph of the 11th edition of the FCA4AI International Workshop (“What can FCA do for Artificial Intelligence?”, FCA4AI 2023, see <https://fca4ai.hse.ru/2023/>). This workshop was co-located with IJCAI 2023, in Macao, S.A.R. China, in August 20 2023. The proceedings are published in the CEUR Workshop Proceedings series (#3489, <http://ceur-ws.org/Vol-3489>).
- Amedeo Napoli is a member of the steering committee of the EKAW Conference (next edition to be held in Amsterdam in Fall 2024). He is also a member of the steering committee of the new Conference “CONCEPTS 2024”, the 1st International Joint Conference on Conceptual Knowledge Structures (<https://concepts2024.uca.es/>). This new event was created at the end of 2023 and joins the three main Conferences about Formal Concept Analysis, namely ICFCA (<https://www.kde.cs.uni-kassel.de/icfca2023/>), CLA (<http://cla.inf.upol.cz/>), and ICCS (<https://iccs-conference.org/>).

10.1.2 Scientific events: selection

Member of the conference program committees The members of Orpailleur are regular PC members of several conferences (e.g., AAI, ECML/PKDD, EKAW, IJCAI, ISMVL, LPIR, EGC) and workshops (e.g., AMLAI, ATA, FCA4AI, IARML, LFA, XKDD).

Reviewer The members of Orpailleur are regular reviewers for international journals (e.g., IJAR, FSS, AMAI, DAM, NEUROCOMPUTING, IEEE-TEVC, Information visualization, etc.)

10.1.3 Journal

Member of the editorial boards

- Miguel Couceiro is member of topical advisory panel of *Axioms*.
- Miguel Couceiro is guest editor of *Order*, *EURO Journal of decision processes*, *Annals of Mathematics and artificial intelligence*, and *Discrete Mathematics and Theoretical Computer Science*.

- Miguel Couceiro is member of the Editorial Board of the *Journal on Multiple-Valued Logic and Soft Computing*, and of the *International Journal on Information Technologies and Security*.

Reviewer - reviewing activities: All members of Orpailleur are involved in journal reviewing activities.

10.1.4 Invited talks

- L. Boudjeloud-Assala, « Humain in the loop, Interactivity and Visualisation », GT RTR-DIAMS, Human in the loop for data mining and machine learning, Orléans – January 2023.

10.1.5 Leadership within the scientific community

- Miguel Couceiro is a member of Management committee of the CNRS GDR RADIA (*Reasoning, Learning, and Decision in Artificial Intelligence*, ex-GDR IA): responsible for corporate training missions.
- Miguel Couceiro is a member of the UL scientific cluster *Automatique, Mathématiques, Informatique et leurs Interactions* (AM2I).
- Miguel Couceiro is member of the *Commission de mention informatique* (CMI) of the UL doctoral school IAEM.
- Miguel Couceiro is an invited member of the Ethics GT of the Hub France IA association⁷ which aims to bring together industrial players in the IA ecosystem to enable the development of proposals and concrete solutions. Among other things, Miguel participated in the drafting and editorial committee of the *livre blanc* “Définition et opérationnalisation des principes de l’IA Ethique”.
- Lydia Boudjeloud-Assala is member of the steering committee and member of the association EGC board (Association Internationale Francophone d’Extraction et de Gestion des Connaissances)

10.1.6 Scientific expertise

- Miguel Couceiro is a member of the scientific board of the *TAILOR Connectivity fund*.
- Miguel Couceiro is a member of the expert panel (math.) of *Research Foundation - Flanders* (FWO)
- The members of Orpailleur are involved in the evaluation of projects for several international and national agencies, e.g. NSERC, FWO, GACR, FWF, ANR, CIFRE, etc.

10.2 Pedagogical responsibilities - Supervision - Juries

10.2.1 Pedagogical responsibilities

- Miguel Couceiro is the responsible of the 2nd year of the NLP M.Sc. at IDMC, UL.
- Miguel Couceiro is the coordinator at UL of the European Erasmus Mundus Master’s program “Language and Communication Technologies” (LCT)
- Miguel Couceiro is the responsible of the international relations at the IDMC, UL.
- Mario Valencia was the academic responsible of the 5nd year of the computer science engineering degree at Polytech Nancy, UL, during the period between September 2022 and July 2023.
- Mario Valencia was the responsible of industry projects for the 4nd and 5nd year of the computer science engineering degree at Polytech Nancy, UL, during the period between September 2022 and July 2023.
- Alain Gély is the responsible of Data Science department of IUT de Metz, UL.

⁷<https://www.hub-franceia.fr/>

- Lydia Boudjeloud-Assala, 3 rd year bachelor's degree coordinator, IUT de Metz (2019-2023)
- Alexandre Blanché, Responsible of master 2 Computer Sciences (UFR MIM) parcours "Systèmes d'Information Décisionnels" (SID)

10.2.2 Supervision

- Safa Alsaidi, Ph.D. student in Computer Science, Inria Paris. "Patient representation learning for analogical reasoning in healthcare" (started: December 2022). Co-supervisors: Adrien Coulet, Miguel Couceiro, Nicolas Garcelon.
- Ambroise Baril, Doctorant en Informatique, LORIA (UMR 7503). "Fine-grained and parameterised complexity of H -colouring problems" (started: September 2021). Co-supervisors: Miguel Couceiro, Victor Lagerkvist⁸.
- Latifa Bouchekif, Ph.D. student in Computer Science, LORIA (UMR 7503). "Approches exploratoire et prédictive pour l'identification des réponses des plantes aux vagues de chaleur récurrentes" (started September 2023). Co-supervisors: Miguel Couceiro, Lydia Boudjeloud-Assala.
- Egor Dudyrev, Ph.D. student in Computer Science, LORIA (UMR 7503). Subject: Making FCA and Pattern Structures Operational, Generic, and Interoperable in the SmartFCA-Platform (started: January 2023). Co-supervisors: Miguel Couceiro, Mehdi Kaytoue.
- Mohammed Fellaji, Ph.D. student in Computer Science, LORIA (UMR 7503). "Réseaux neuronaux bayésiens dynamiques" (started: January 2022). Co-supervisors: Miguel Couceiro, Frédéric Pennerath.
- Lucas Jarnac, Ph.D. student in Computer Science, Orange (CIFRE contract). "Réconcilier des connaissances incertaines pour améliorer un graphe de connaissances" (started: February 2023). Co-supervisors: Yoann Chabot, Miguel Couceiro.
- Esteban Marquer, Doctorant en Informatique, LORIA (UMR 7503). "Reasoning over Data: Analogy-based and transfer learning to improve Machine Learning" (started: October 2020). Co-supervisors: Miguel Couceiro, Alain Gély.
- Morgan Ruiz-Huidobro, Ph.D. student in Computer Science, Happiso (CIFRE contract). "Exploiting explanations and fairness of neural approaches for a decision support system" (started: April 2023). Co-supervisors: Miguel Couceiro, Bernadetta Addis.
- Judy Akl, Ph.D. student in Computer Science, (LCOMS EA 7306) "Real-time assets tracking and monitoring for industry 4.0: Towards a new IoT solution service based on Bluetooth devices and AI for data fusion" (started: February 2023). Co-supervisors: Fahed Abdallah (LCOMS), Alexandre Blanché.

10.2.3 Juries

- Participation in the PhD jury of RutaBinkyte-Sadauskiene, *Advancing Ethical AI: Methods for fairness enhancement leveraging on causality and under privacy constraints*, Inria/LIX/Institut Polytechnique de Paris (UMR 7161), December 2023. Miguel Couceiro, member.
- Participation in the PhD jury of Thomas Ranvier, *Deep Learning on Incomplete and Multi-Source Data: Application to Cancer Immunotherapy*, Université Lyon 1, December 2003. Lydia Boudjeloud-Assala, reviewer.
- Participation in the PhD jury of Noémie Gonnier, *CxSOM : vers une architecture non hiérarchique de cartes auto-organisatrices Méthodes et outils d'analyse*, Université de Lorraine, June 2023. Lydia Boudjeloud-Assala, member.

⁸Department of Computer and Information Science, Linköping University, Sweden

- Participation in the PhD jury of Nassim Haddam, *Reinforcement learning for lighting energy control in a connected building*, Université de Versailles-SQ, March 2023. Lydia Boudjeloud-Assala, reviewer.
- Amedeo Napoli participated in December 2023 in the HDR Defense Committee of Nizar Messai at the University of Tours. The title of the document is *Supporting Service Interactions with Semantics, Machine Learning, and Blockchain*, and Amedeo Napoli served as the president of the committee.

10.3 Popularization

10.3.1 Interventions

- Mars 2023 : Intervention France Bleu Lorraine Nord, “côté expert : Chat GPT, c’est quoi ? Comment fonctionne l’Intelligence Artificielle (IA) ?” by Lydia Boudjeloud-Assala

10.3.2 Education

- Lydia Boudjeloud-Assala a accueilli 2 lycéens dans son cours dans le cadre de UJALU « Un jour à l’université »
- Miguel Couceiro a accueilli Lilian Tollé (Collège Croix de Metz) au sein d’Orpailleur pour son “Stage Découverte”.

11 Scientific production

11.1 Major publications

- [1] P. Andrieu, S. Cohen-Boulakia, M. Couceiro, A. Denise and A. Pierrot. ‘A unifying rank aggregation framework to suitably and efficiently aggregate any kind of rankings’. In: *International Journal of Approximate Reasoning* 162 (22nd Sept. 2023), p. 109035. DOI: [10.1016/j.ijar.2023.109035](https://doi.org/10.1016/j.ijar.2023.109035). URL: <https://inria.hal.science/hal-04227695>.
- [2] G. D. Pelegrina, M. Couceiro and L. T. Duarte. ‘A statistical approach to detect disparity prone features in a group fairness setting’. In: *AI and Ethics* (Nov. 2023). DOI: [10.48550/arXiv.2305.06994](https://doi.org/10.48550/arXiv.2305.06994). URL: <https://inria.hal.science/hal-04096649>.
- [3] L. Jarnac, M. Couceiro and P. Monnin. ‘Relevant Entity Selection: Knowledge Graph Bootstrapping via Zero-Shot Analogical Pruning’. In: *CIKM ’23: The 32nd ACM International Conference on Information and Knowledge Management*. Birmingham (UK): ACM, 21st Oct. 2023, pp. 934–944. DOI: [10.1145/3583780.3615030](https://doi.org/10.1145/3583780.3615030). URL: <https://inria.hal.science/hal-04254979>.

11.2 Publications of the year

International journals

- [4] N. Abbas, A. Bazin, J. David and A. Napoli. ‘Discovery of link keys in resource description framework datasets based on pattern structures’. In: *International Journal of Approximate Reasoning* 161 (Oct. 2023), p. 108978. DOI: [10.1016/j.ijar.2023.108978](https://doi.org/10.1016/j.ijar.2023.108978). URL: <https://hal.science/hal-04368386>.
- [5] G. Alves, F. Bernier, M. Couceiro, K. Makhoul, C. Palamidessi and S. Zhioua. ‘Survey on Fairness Notions and Related Tensions’. In: *EURO journal on decision processes* (2023). DOI: [10.1016/j.ejdp.2023.100033](https://doi.org/10.1016/j.ejdp.2023.100033). URL: <https://hal.science/hal-03484009>.
- [6] M. Couceiro and E. Lehtonen. ‘Galois theory for analogical classifiers’. In: *Annals of Mathematics and Artificial Intelligence* (2023). DOI: [10.1007/s10472-023-09833-6](https://doi.org/10.1007/s10472-023-09833-6). URL: <https://inria.hal.science/hal-03665625>.

- [7] F. Namer, N. Hathout, D. Amiot, L. Barque, O. Bonami, G. Boyé, B. Calderone, J. Cattini, G. Dal, A. Delaporte, G. Duboisdindien, A. Falaise, N. Grabar, P. Haas, F. Henry, M. Huguin, N. Juniarta, L. Liégeois, S. Lignon, L. Macchi, G. Manucharian, C. Masson, F. Montermini, N. Okinina, F. Sajous, D. Sanacore, M. T. Tran, J. Thuilier, Y. Toussaint and D. Tribout. ‘Démonette-2, a derivational database for French with broad lexical coverage and fine-grained morphological descriptions’. In: *Lexique* 33 (22nd Dec. 2023), pp. 6–40. DOI: [10.54563/lexique.1242](https://doi.org/10.54563/lexique.1242). URL: <https://hal.science/hal-04363595>.
- [8] M. Couceiro and E. Lehtonen. ‘Stability of Boolean function classes with respect to clones of linear functions’. In: *Order* (). DOI: [10.1007/s11083-022-09596-5](https://doi.org/10.1007/s11083-022-09596-5). URL: <https://hal.science/hal-03119362>.

International peer-reviewed conferences

- [9] F. Badra, M.-J. Lesot, E. Marquer and M. Couceiro. ‘Some Perspectives on Similarity Learning for Case-Based Reasoning and Analogical Transfer’. In: *Proceedings of the 2nd Workshop Interactions between Analogical Reasoning and Machine Learning, IARML@IJCAI’2023*. IARML@IJCAI’2023: Workshop on the Interactions between Analogical Reasoning and Machine Learning, at IJCAI’2023. Vol. 3492. CEUR-WS.org. Macao, China: CEUR-WS.org, 23rd Sept. 2023, pp. 16–29. URL: <https://inria.hal.science/hal-04208969>.
- [10] J. Baixeries, V. Codocedo, M. Kaytoue and A. Napoli. ‘Three Views on Dependency Covers from an FCA Perspective’. In: *Proceedings of the 17th International Conference on Formal Concept Analysis (ICFCA, Lecture Notes in Computer Science 13934)*. 17th International Conference on Formal Concept Analysis (ICFCA). Vol. Lecture Notes in Computer Science. Lecture Notes in Computer Science 13934 13934. Kassel university, Germany: Springer, 2023, pages 78–94. DOI: [10.1007/978-3-031-35949-1_6](https://doi.org/10.1007/978-3-031-35949-1_6). URL: <https://inria.hal.science/hal-04370042>.
- [11] E. Dudyrev, S. O. Kuznetsov and A. Napoli. ‘Constructing Decision Quivers’. In: *CEUR Workshop Proceedings, FCA4AI 2023 What can FCA do for Artificial Intelligence 2023*. "What can FCA do for Artificial Intelligence?" (FCA4AI 2023). Vol. 3489. Macao SAR, Macao SAR China, 23rd Sept. 2023, pp. 69–80. URL: <https://hal.science/hal-04364229>.
- [12] F. Gaschi, X. Fontaine, P. Rastin and Y. Toussaint. ‘Multilingual Clinical NER: Translation or Cross-lingual Transfer?’ In: *Proceedings of the 5th Clinical Natural Language Processing Workshop*. 5th Clinical Natural Language Processing Workshop. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 289–311. DOI: [10.18653/v1/2023.clinicalnlp-1.34](https://doi.org/10.18653/v1/2023.clinicalnlp-1.34). URL: <https://hal.science/hal-04193182>.
- [13] E. Marquer, F. Badra, M.-J. Lesot, M. Couceiro and D. Leake. ‘Less is Better: An Energy-Based Approach to Case Base Competence’. In: *Proceedings of the Workshops at the 31st International Conference on Case-Based Reasoning (ICCB-WS 2023)co-located with the 31st International Conference on Case-Based Reasoning (ICCB 2023)*. ICCBR ATA 2023: Workshop on Analogies: From Theory to Applications “AR & CBR Tools for Metric and Representation Learning at ICCBR2023. Vol. 3438. Aberdeen, United Kingdom: CEUR, 17th July 2023, pp. 27–42. URL: <https://inria.hal.science/hal-04184905>.

Conferences without proceedings

- [14] F. Gaschi, P. Cerda, P. Rastin and Y. Toussaint. ‘Exploring the Relationship between Alignment and Cross-lingual Transfer in Multilingual Transformers’. In: *Findings of the Association for Computational Linguistics: ACL 2023*, Findings of the Association for Computational Linguistics: ACL 2023. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 3020–3042. DOI: [10.18653/v1/2023.findings-acl.189](https://doi.org/10.18653/v1/2023.findings-acl.189). URL: <https://hal.science/hal-04193179>.
- [15] L. Zanella and Y. Toussaint. ‘Adding Linguistic Information to Transformer Models Improves Biomedical Event Detection?’ In: 1st Symposium on Challenges for Natural Language Processing (CNLPSTM23). Varsovie, Poland, 17th Sept. 2023. URL: <https://hal.science/hal-04197974>.

- [16] L. Zanella and Y. Toussaint. ‘How Much do Knowledge Graphs Impact Transformer Models for Extracting Biomedical Events?’ In: The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks. Vol. The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks. Toronto, Canada, 9th July 2023, 145–155. DOI: [10.18653/v1/2023.bionlp-1.12](https://doi.org/10.18653/v1/2023.bionlp-1.12). URL: <https://hal.science/hal-04192212>.

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

- [17] F. Badra, M. Couceiro, E. Marquer and P. Monnin, eds. *Proceedings of the 2nd Workshop Analogies: from Theory to Applications (ATA@ICCBR 2023): AR & CBR Tools for Metric and Representation Learning*. 31st International Conference on Case-Based Reasoning (ICCBR 2023). Vol. 3438. ICCBR 2023 Workshop Proceedings. CEUR-WS.org, 8th July 2023, pp. 2–53. URL: <https://inria.hal.science/hal-04392016>.
- [18] M. Couceiro, E. Marquer, P.-A. Murena and P. Monnin, eds. *Workshop Analogies: from Theory to Applications (ATA@ICCBR 2022)*. Workshop Analogies: from Theory to Applications (ATA@ICCBR 2022). Vol. 3389. Workshop Proceedings of the 30th International Conference on Case-Based Reasoning co-located with the 30th International Conference on Case-Based Reasoning (ICCBR 2022). CEUR-WS.org, 11th May 2023, pp. 3–103. URL: <https://inria.hal.science/hal-04392022>.
- [19] M. Couceiro, P.-A. Murena and S. Afantenos, eds. *Proceedings of the 2nd Workshop on the Interactions between Analogical Reasoning and Machine Learning*. Workshop on the Interactions between Analogical Reasoning and Machine Learning (IARML@IJCAI 2023). Vol. 3492. CEUR-WS.org, 23rd Sept. 2023, p. 57. URL: <https://inria.hal.science/hal-04391981>.
- [20] S. O. Kuznetsov, A. Napoli and S. Rudolph, eds. *Proceedings of the 10th International Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI 2023)*. International Conference on Artificial Intelligence (IJCAI 2023), Macao, S.A.R. China. Vol. CEUR Proceedings (<http://ceur-ws.org/Vol-3489>). 3489. CEUR-WS.org, 20th Aug. 2023, p. 80. URL: <https://inria.hal.science/hal-04370014>.

Doctoral dissertations and habilitation theses

- [21] C. Theobald. ‘Bayesian Deep Learning for Mining and Analyzing Astronomical Data’. Université de Lorraine, 6th June 2023. URL: <https://hal.univ-lorraine.fr/tel-04206212>.
- [22] G. Zervakis. ‘Enriching large language models with semantic lexicons and analogies’. Université de Lorraine, 8th Mar. 2023. URL: <https://theses.hal.science/tel-04138899>.

Reports & preprints

- [23] A. Baril, A. Castillon and N. Oijid. *On the parameterized complexity of non-hereditary relaxations of clique*. Liris; Loria; CRISTAL, 2023. URL: <https://hal.science/hal-04036849>.
- [24] A. Buzmakov, E. Dudyrev, S. O. Kuznetsov, T. Makhalova and A. Napoli. *Data complexity: An FCA-based approach*. 20th Feb. 2023. URL: <https://hal.science/hal-03985980>.
- [25] A. Buzmakov, S. O. Kuznetsov, T. Makhalova, A. Napoli and E. Dudyrev. *Data complexity: An FCA-based approach*. 30th Jan. 2023. URL: <https://hal.science/hal-03970678>.
- [26] E. Dudyrev, S. O. Kuznetsov and A. Napoli. *Description Quivers for Compact Representation of Concept Lattices and Ensembles of Decision Trees*. 22nd May 2023. URL: <https://hal.science/hal-04092794>.
- [27] A. Kulkarni, A. Kulkarni, M. Couceiro and H. Aldarmaki. *Adapting the adapters for code-switching in multilingual ASR*. 2023. DOI: [10.48550/arXiv.2310.07423](https://doi.org/10.48550/arXiv.2310.07423). URL: <https://hal.science/hal-04238811>.
- [28] E. Marquer and M. Couceiro. *Solving morphological analogies: from retrieval to generation*. 30th Mar. 2023. URL: <https://hal.science/hal-04056908>.

- [29] M. R. Qureshi, L. Galárraga and M. Couceiro. *A reinforcement learning approach to mitigating stereotypical biases in language models*. 15th Nov. 2023. URL: <https://inria.hal.science/hal-04426115>.
- [30] A. Baril, M. Couceiro and V. Lagerkvist. ‘Linear Bounds between Component Twin-Width and Clique-Width with Algorithmic Applications to Counting Graph Colorings’. working paper or preprint. Dec. 2023. URL: <https://hal.science/hal-04142719>.

11.3 Cited publications

- [31] G. Alves, M. Amblard, F. Bernier, M. Couceiro and A. Napoli. ‘Reducing Unintended Bias of ML Models on Tabular and Textual Data’. In: *DSAA 2021 - 8th IEEE International Conference on Data Science and Advanced Analytics*. Porto (virtual event), Portugal, Oct. 2021, pp. 1–10. URL: <https://hal.science/hal-03312797>.
- [32] A. Baril, M. Couceiro and V. Lagerkvist. ‘New Perspectives on Semiring Applications to Dynamic Programming’. working paper or preprint. Feb. 2024. URL: <https://hal.science/hal-04434071>.
- [33] E. M. Bender, T. Gebru, A. McMillan-Major and S. Shmitchell. ‘On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?’ In: *FACCT ’21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021*. Ed. by M. C. Elish, W. Isaac and R. S. Zemel. ACM, 2021, pp. 610–623.
- [34] V. Bhargava, M. Couceiro and A. Napoli. ‘LimeOut: An Ensemble Approach To Improve Process Fairness’. In: *ECML-PKDD International Workshop on eXplainable Knowledge Discovery in Data Mining (XKDD 2020)*. Vol. 1323. Communications in Computer and Information Science. Ghent, Belgium: Springer, Sept. 2020, pp. 475–491. URL: <https://hal.science/hal-02979233>.
- [35] P. Delobelle, E. K. Tokpo, T. Calders and B. Berendt. ‘Measuring Fairness with Biased Rulers: A Comparative Study on Bias Metrics for Pre-trained Language Models’. In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022*. Ed. by M. Carpuat, M. de Marneffe and I. V. M. Ruíz. Association for Computational Linguistics, 2022, pp. 1693–1706.
- [36] R. Galian, T. Hamm, V. Korchemna, K. Okrasa and K. Simonov. ‘The Fine-Grained Complexity of Graph Homomorphism Parameterized by Clique-Width’. In: *49th International Colloquium on Automata, Languages, and Programming, ICALP 2022, July 4-8, 2022, Paris, France*. Ed. by M. Bojanczyk, E. Merelli and D. P. Woodruff. Vol. 229. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022, 66:1–66:20. DOI: [10.4230/LIPICS.ICALP.2022.66](https://doi.org/10.4230/LIPICS.ICALP.2022.66). URL: <https://doi.org/10.4230/LIPICS.ICALP.2022.66>.
- [37] Y. Gao, S. Zeng, X. Xu, H. Li, S. Yao, K. Song, X. Li, L. Chen, J. Tang, H. Xing et al. ‘Deep learning-enabled pelvic ultrasound images for accurate diagnosis of ovarian cancer in China: a retrospective, multicentre, diagnostic study.’ In: *The Lancet. Digital health* 4 3 (2022), e179–e187. URL: <https://api.semanticscholar.org/CorpusID:247089260>.
- [38] S. Goldfarb-Tarrant, R. Marchant, R. M. Sánchez, M. Pandya and A. Lopez. ‘Intrinsic Bias Metrics Do Not Correlate with Application Bias’. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*. Ed. by C. Zong, F. Xia, W. Li and R. Navigli. Association for Computational Linguistics, 2021, pp. 1926–1940.
- [39] P. Jensen, L. Jensen and S. Brunak. ‘Mining electronic health records: towards better research applications and clinical care.’ English. In: *Nature Reviews. Genetics* 13.6 (2012), pp. 395–405. DOI: [10.1038/nrg3208](https://doi.org/10.1038/nrg3208).
- [40] I. Landi, B. S. Glicksberg, H. Lee, S. T. Cherng, G. Landi, M. Danieletto, J. T. Dudley, C. Furlanello and R. Miotto. ‘Deep representation learning of electronic health records to unlock patient stratification at scale’. In: *npj Digit. Medicine* 3 (2020). DOI: [10.1038/s41746-020-0301-z](https://doi.org/10.1038/s41746-020-0301-z). URL: <https://doi.org/10.1038/s41746-020-0301-z>.

- [41] T. Li, D. Khashabi, T. Khot, A. Sabharwal and V. Srikumar. ‘UNQOVERing Stereotyping Biases via Underspecified Questions’. In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, Nov. 2020, pp. 3475–3489. DOI: [10.18653/v1/2020.findings-emnlp.311](https://doi.org/10.18653/v1/2020.findings-emnlp.311). URL: <https://aclanthology.org/2020.findings-emnlp.311>.
- [42] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova and C. Zhong. ‘Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges’. In: *CoRR abs/2103.11251* (2021). arXiv: [2103.11251](https://arxiv.org/abs/2103.11251). URL: <https://arxiv.org/abs/2103.11251>.
- [43] A. Schuler, A. Callahan, K. Jung and N. H. Shah. ‘Performing an Informatics Consult: Methods and Challenges’. In: *Journal of the American College of Radiology* 15.3, Part B (2018). Data Science: Big Data Machine Learning and Artificial Intelligence, pp. 563–568. DOI: <https://doi.org/10.1016/j.jacr.2017.12.023>. URL: <https://www.sciencedirect.com/science/article/pii/S154614401731668X>.
- [44] Y. Si, J. Du, Z. Li, X. Jiang, T. A. Miller, F. Wang, W. J. Zheng and K. Roberts. ‘Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review’. In: *J. Biomed. Informatics* 115 (2021), p. 103671. DOI: [10.1016/J.JBI.2020.103671](https://doi.org/10.1016/J.JBI.2020.103671). URL: <https://doi.org/10.1016/j.jbi.2020.103671>.
- [45] C. Theobald, F. Pennerath, B. Conan-Guez, M. Couceiro and A. Napoli. *Clarity: an improved gradient method for producing quality visual counterfactual explanations*. 2022. arXiv: [2211.15370](https://arxiv.org/abs/2211.15370) [cs.CV].
- [46] G. Zervakis, E. Vincent, M. Couceiro, M. Schoenauer and E. Marquer. ‘An analogy based approach for solving target sense verification’. In: *NLPIR 2022 - 6th International Conference on Natural Language Processing and Information Retrieval*. Bangkok, Thailand, Dec. 2022. URL: <https://inria.hal.science/hal-03792071>.