

2025 Activity Report

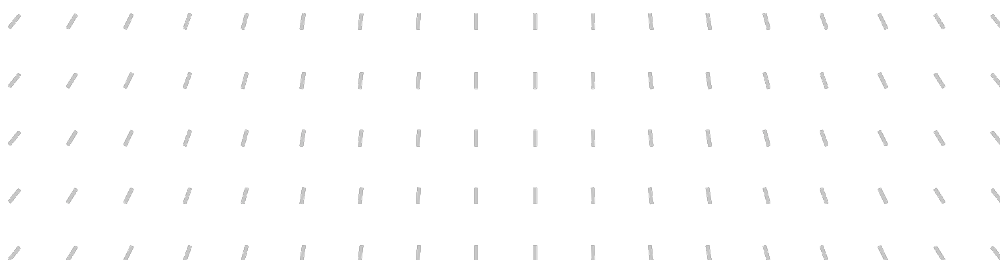
RESEARCH CENTRE: Inria Centre at the University of Lille
IN PARTNERSHIP WITH: CNRS, Université de Lille

Project-Team

MAGNET

Machine Learning in Information Networks

In collaboration with Centre de Recherche en Informatique, Signal et Automatique
de Lille



Project-Team MAGNET

Creation of the Project-Team: 2016 May 01

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

Keywords

Computer sciences and digital sciences

- A3.1.11. – Structured data
- A3.4. – Machine learning and statistics
- A5.7.2. – Music
- A5.7.3. – Speech
- A5.8. – Natural language processing
- A9.1. – Knowledge
- A9.2. – Machine learning
- A9.4. – Natural language processing
- A9.7. – AI algorithmics
- A9.8. – Reasoning
- A9.9. – Distributed AI, Multi-agent
- A9.10. – Hybrid approaches for AI
- A9.11. – Generative AI
- A9.13. – Agentic AI
- A9.14. – Evaluation of AI models
- A9.17. – Cybersecurity and AI

Other research topics and application domains

- B2. – Digital health
- B9.5.1. – Computer science
- B9.5.6. – Data science
- B9.6.1. – Psychology
- B9.6.8. – Linguistics
- B9.6.10. – Digital humanities
- B9.9. – Ethics
- B9.10. – Privacy

Contents

Project-Team MAGNET	1
1 Team members, visitors, external collaborators	5
2 Overall objectives	6
3 Research program	7
4 Application domains	9
5 Social and environmental responsibility	9
5.1 Footprint of research activities	9
5.2 Impact of research results	10
6 Highlights of the year	10
7 Latest software developments, platforms, open data	10
7.1 Latest software developments	10
7.1.1 CoRTeX	10
7.1.2 Mangoes	10
7.1.3 metric-learn	11
7.1.4 MyLocalInfo	11
7.1.5 decllearn	11
7.1.6 fairgrad	12
7.1.7 tasksource	12
7.1.8 Voice Transformer 2	12
7.2 Open data	13
8 New results	13
8.1 Natural Language Processing	13
8.2 Privacy and NLP	17
8.3 Music and NLP	18
8.4 Speech and Privacy	19
8.5 Security and Privacy	20
8.6 Fairness	23
8.7 Federated and Decentralized Learning	24
8.8 Conformal Prediction for Trustworthy Machine Learning	25
9 Bilateral contracts and grants with industry	26
9.1 Bilateral contracts with industry	26
10 Partnerships and cooperations	26
10.1 International research visitors	26
10.1.1 Visits of international scientists	26
10.1.2 Visits to international teams	27
10.2 European initiatives	27
10.2.1 Horizon Europe	27
10.3 National initiatives	29
10.3.1 ANR PMR (2020-2025)	29
10.3.2 FedMalin. INRIA Defi (2021-2026)	29
10.3.3 COMANCHE: Computational Models of Lexical Meaning and Change. INRIA Action Exploratoire (2022-2026)	30
10.3.4 IPoP, Projet interdisciplinaire sur la protection des données personnelles, PEPR Cybersécurité (2022-2028).	30

10.3.5	SSF-ML-DH, PEPR Santé Numérique (2022-2028)	31
10.3.6	CAPS'UL (2023-2028)	31
10.3.7	ANR-JCJC FaCTor: Fairness Constraints and Guarantees for Trustworthy Machine Learning (2023-2027)	32
10.3.8	REDEEM: Resilient, Decentralized and Privacy-Preserving Machine Learning, PEPR IA (2022-2028)	32
10.3.9	ANR-JCJC Adada: Adada: Adaptive Datasets for Enhancing Reasoning in Large Language Models (2024-2028)	32
10.3.10	ANR Melissa: METHodological contributions in statistical LEARNING InSPIred by SurfACE engineering (2025-2029)	33
10.4	Regional initiatives	33
10.4.1	Cross Disciplinary Project (CDP) Prime Next Gen: NEXT-GENERATION PReCIson medicine in Inflammatory and MEtabolic diseases (2026-2030)	33
10.4.2	Cross Disciplinary Project (CDP) LOOP: closed Loop neurOTEchnologies: from sensORS to aPplications (2026-2030)	34
11	Dissemination	34
11.1	Promoting scientific activities	34
11.1.1	Scientific events: selection	34
11.1.2	Journal	34
11.1.3	Invited talks	35
11.1.4	Leadership within the scientific community	35
11.1.5	Scientific expertise	35
11.1.6	Research administration	35
11.2	Teaching - Supervision - Juries - Educational and pedagogical outreach	36
11.2.1	Teaching	36
11.2.2	Supervision	36
11.2.3	Juries	38
11.2.4	Educational and pedagogical outreach	38
11.3	Popularization	38
11.3.1	Participation in Live events	38
11.3.2	Others science outreach relevant activities	38
12	Scientific production	38
12.1	Major publications	38
12.2	Publications of the year	40

1 Team members, visitors, external collaborators

Research Scientists

- Pascal Denis [INRIA, Researcher]
- Raouf Kerkouche [INRIA, Researcher, from Oct 2025]
- Batiste Le Bars [INRIA, ISFP]
- Michael Perrot [INRIA, ISFP]
- Jan Ramon [INRIA, Senior Researcher, HDR]
- Damien Sileo [INRIA, ISFP]

Faculty Members

- Marc Tommasi [Team leader, UNIV LILLE, Professor, HDR]
- Angèle Brunelliere [UNIV LILLE, Professor Delegation, until Aug 2025]
- Remi Gilleron [UNIV. LILLE, Emeritus, HDR]
- Mikaela Keller [UNIV LILLE, Associate Professor, from Sep 2025]
- Mikaela Keller [UNIV LILLE, Associate Professor Delegation, until Aug 2025]

Post-Doctoral Fellows

- Arnaud Descours [INRIA, Post-Doctoral Fellow, until Aug 2025]
- Luis Eduardo Lugo Martinez [INRIA, Post-Doctoral Fellow]

PhD Students

- Paul Andrey [INRIA]
- Antoine Barczewski [INRIA, until Oct 2025]
- Mouad Blej [INRIA]
- Nassim Boudjenah [INRIA, from Dec 2025]
- Edwige Cyffers [UNIV LILLE, until Mar 2025]
- Marc Damie [UNIV. TWENTE]
- Jean Dufraiche [INRIA]
- Brahim Erraji [INRIA]
- Dimitri Kachler [INRIA, from Nov 2025]
- Aleksei Korneev [UNIV LILLE]
- Dinh-Viet-Toan Le [UNIV LILLE, until Oct 2025]
- Bastien Lietard [INRIA]
- Gabriel Loiseau [HORNET SECURITY, CIFRE]
- Aymane Moataz [INRIA, until Jun 2025]

- Clement Pierquin [CRAFT.AI, CIFRE]
- Aurelien Said Housseini [INRIA, until May 2025]
- Quentin Sinh [INRIA]
- Shreya Venugopal [INRIA]

Technical Staff

- Amer Alzein [INRIA, Engineer, from Jun 2025 until Sep 2025]
- Jules Boulet [INRIA, Engineer, until May 2025]
- Baptiste Cottier [INRIA, Engineer]
- Simon Decomble [INRIA, Engineer, from Apr 2025]
- Leonard Deroose [INRIA, Engineer]
- Zakaria El Bouchouari [INRIA, Engineer, from Sep 2025]
- Younes Ikli [INRIA, Engineer]
- Valentin Lacombe [INRIA, Engineer, from Feb 2025]
- Alexandre Louvet [INRIA, Engineer, from Sep 2025]
- Victor Roussanaly [INRIA, Engineer, from Sep 2025]
- Elina Thibeau-Sutre [INRIA, Engineer, until Jun 2025]
- Jules Yvon [INRIA, Engineer]

Interns and Apprentices

- Thomas Bobille [INRIA, Intern, from Apr 2025 until Aug 2025]
- Yassine Oj [INRIA, Intern, from Apr 2025 until May 2025]
- Valentin Quesnel-Dumont [INRIA, Intern, from Apr 2025 until Aug 2025]
- Mohamed El Amine Serradj [INRIA, Intern, from Apr 2025 until Aug 2025]

Administrative Assistants

- Nathalie Bonte [INRIA, from Jun 2025]
- Aurore Dalle [INRIA, until May 2025]

2 Overall objectives

The main objective of MAGNET is to develop original machine learning methods for networked data. We consider information networks in which the data consist of feature vectors or texts. We model such networks as graphs wherein nodes correspond to entities (documents, spans of text, users, datasets, learners etc.) and edges correspond to relations between entities (similarity, answer, co-authoring, friendship etc.). In *Mining and Learning in Graphs*, our main research goal is to efficiently search for the best hidden graph structure to be generated for solving a given learning task which exploits the relationships between entities. In *Machine Learning for Natural Language Processing* the objective is to go beyond vectorial classification to solve tasks like coreference resolution and entity linking, temporal structure prediction, and discourse parsing. In

Decentralized Machine Learning we address the problem of learning in a private, fair and energy efficient way when data are naturally distributed in a network.

The challenges are the dimensionality of the input space, possibly the dimensionality of the output space, the high level of dependencies between the data, the inherent ambiguity of textual data and the limited amount of human labeling. We are interested in making machine learning approaches more acceptable to society. Privacy, sobriety and fairness are important issues that pertain to this research line, and we are interested in the empowerment of end users in the machine learning processes.

3 Research program

The research program of MAGNET is structured along three main axes.

Axis 1: Mining and Learning in Graphs This axis is the backbone of the team. Most of the techniques and algorithms developed in this axis are known by the team members and have impact on the two other axes. We address the following questions and objectives:

How to adaptively build graphs with respect to the given tasks? We study adaptive graph construction along several directions. The first one is to learn the best similarity measure for the graph construction. The second one is to combine different views over the data in the graph construction and learn good representations. We also study weak forms of supervision like comparisons.

How to design methods able to achieve a good trade-off between predictive accuracy and computational complexity? We develop new algorithms for efficient graph-based learning (for instance node prediction or link prediction). In order to deal with scalability issues, our approach is based on optimization, graph sparsification techniques and graph sampling methods.

How to find patterns in graphs based on efficient computations of some statistics? We develop graph mining algorithms and statistics in the context of correlated data.

Axis 2: Machine Learning for Natural Language Processing In this axis, we address the general question that relates graph-based learning and Natural Language Processing (NLP): *How to go beyond vectorial classification models in NLP tasks?* We study the combination of learning representation, structured prediction and graph-based learning methods. Data sobriety and fairness are major constraints we want to deal with. The targeted NLP tasks are coreference resolution and entity linking, temporal structure prediction, and discourse parsing.

Axis 3: Decentralized Machine Learning and Privacy In this axis, we study *How to design private by design machine learning algorithms?* Taking as an opportunity the fact that data collection is now decentralized on smart devices, we propose alternatives to large data centers where data are gathered by developing collaborative and personalized learning.

Contrary to many machine learning approaches where data points and tasks are considered in isolation, we think that a key point of this research is to be able to leverage the relationships between data and learning objectives. Therefore, using graphs as an abstraction of information networks is a major playground for MAGNET. Research related to graph data is a transversal axis, describing a layer of work supporting two other axes on Natural Language Processing and decentralized learning. The machine learning and mining in graphs communities have evolved, for instance taking into account data streams, dynamics but maybe more importantly, focusing on deep learning. Deep neural nets are here to stay, and they are useful tools to tackle difficult problems so we embrace them at different places in the three axes.

MAGNET conducts research along the three axes described above but will put more emphasis on social issues of machine learning. In the context of the recent deployment of artificial intelligence into our daily lives, we are interested in making machine learning approaches more acceptable to society. Privacy, sobriety and fairness are important issues that pertain to this research line, but more generally we are interested in the empowerment of end users in the machine learning processes. Reducing the need of one central authority and pushing more the data processing on the user side, that is decentralization, also participates to this effort. Reducing resources means reducing costs and energy and contributes to building more accessible technologies for companies and users. By considering learning tasks in a more personalized way, but

increasing collaboration, we think that we can design solutions that work in low resources regime, with less data or supervision.

In MAGNET we emphasize a different approach than blindly brute-forcing tasks with loads of data. Applications to social sciences for instance have different needs and constraints that motivate data sobriety, fairness and privacy. We are interested in weaker supervision, by leveraging structural properties described in graphs of data, relying on transfer and multi-task learning when faced with graphs of tasks and users. Algorithmic and statistical challenges related to the graph structure of the data still contain open questions. On the statistical side, examples are to take dependencies into account, for instance to compute a mean, to reduce the need of sampling by exploiting known correlations. For the algorithmic point of view, going beyond unlabeled undirected graphs, in particular considering attributed graphs containing text or other information and addressing the case of distributed graphs while maintaining formal guarantees are getting more attention.

In the second axis devoted to NLP, we focus our research on graph-based and representation learning into several directions, all aiming at learning *richer, more robust, and more transferable linguistic representations*. This research program will attempt to bring about strong cross-fertilizations with the other axes, addressing problems in graph, privacy and fairness and making links with decentralized learning. At the intersection between graph-based and representation learning, we will first develop graph embedding algorithms for deriving linguistic representations which are able to capture higher-level semantic and world-knowledge information which eludes strictly distributional models. As an initial step, we envision leveraging pre-existing ontologies (e.g., WordNet, DBpedia), from which one can easily derive interesting similarity graphs between words or noun phrases. We also plan to investigate innovative ways of articulating graph-based semi-supervised learning algorithms and word embedding techniques. A second direction involves learning representations that are more robust to bias, privacy attacks and adversarial examples. Thus, we intend to leverage recent adversarial training strategies, in which an adversary attempts to recover sensitive attributes (e.g., gender, race) from the learned representations, to be able to neutralize bias or to remove sensitive features. An application domain for this line of research is for instance speech data. The study of learning private representation with its link to fairness in the decentralized setting is another important research topic for the team. In this context of fairness, we also intend to develop similar algorithms for detecting slants, and ultimately for generating de-biased or “re-biased” versions of text embeddings. An illustration is on political slant in written texts (e.g., political speeches and manifestos). Thirdly, we intend to learn linguistic representations that can transfer more easily across languages and domains, in particular in the context of structured prediction problems for low-resource languages. For instance, we first propose to jointly learn model parameters for each language (and/or domains) in a multi-task setting, and leverage a (pre-existing or learned) graph encoding structural similarities between languages (and/or domains). This type of approach would nicely tie in with our previous work on multilingual dependency parsing and on learning personalized models. Furthermore, we will also study how to combine and adapt some neural architectures recently introduced for sequence-to-sequence problems in order to enable transfer of language representations.

In terms of technological transfer, we maintain collaborations with researchers in the humanities and the social sciences, helping them to leverage state-of-the-art NLP techniques to develop new insights to their research by extracting relevant information from large amounts of texts.

The third axis is on distributed and decentralized learning and privacy preserving machine learning. Recent years have seen the evolution of information systems towards ubiquitous computing, smart objects and applications fueled by artificial intelligence. Data are collected on smart devices like smartphones, watches, home devices etc. They include texts, locations, social relationships. Many sensitive data —race, gender, health conditions, tastes etc— can be inferred. Others are just recorded like activities, social relationships but also biometric data like voice and measurements from sensor data. The main tendency is to transfer data into central servers mostly owned by a few tier parties. The situation generates high privacy risks for the users for many reasons: loss of data control, unique entry point for data access, unsolicited data usage etc. But it also increases monopolistic situations and tends to develop oversized infrastructures. The centralized paradigm also has limits when data are too huge such as in the case of multiple videos and sensor data collected for autonomous driving. Partially or fully decentralized systems provide an alternative, to emphasis data exploitation rather than data sharing. For MAGNET, they are source of many new research directions in machine learning at two scales: at the algorithmic level and at a systemic level.

At the algorithmic level the question is to develop new privacy preserving algorithms in the context of

decentralized systems. In this context, data remains where it has been collected and learning or statistical queries are processed at the local level. An important question we study is to take into account and measure the impact of collaboration. We also aim at developing methods in the online setting where data arrives continuously or participants join and leave the collaboration network. The granularity of exchanges, the communication cost and the dynamic scenarios, are also studied. On the privacy side, decentralization is not sufficient to establish privacy guarantees because learned models together with the dynamics of collaborative learning may reveal private training data if the models are published or if the communications are observed. But, although it has not been yet well established, decentralization can naturally increase privacy-utility ratio. A direction of research is to formally prove the privacy gain when randomized decentralized protocols are used during learning. In some situations, for instance when part of the data is not sensitive or when trusted servers can be used, a combination between a fully decentralized and a centralized approach is very relevant. In this setting, the question is to find a good trade-off between local versus global computations.

At the systemic layer, in MAGNET we feel that there is a need for research on a global and holistic level, that is to consider full processes involving learning, interacting, predicting, reasoning, repeating etc. rather than studying the privacy of isolated learning algorithms. Our objective is to design languages for describing processes (workflows), data (database schema, background knowledge), population statistics, privacy properties of algorithms, privacy requirements and other relevant information. This is fully aligned with recent trends that aim at giving to statistical learning a more higher level of formal specifications and illustrates our objective for more acceptable and transparent machine learning. We also work towards more robust privacy-friendly systems, being able to handle a wider range of malicious behavior such as collusion to obtain information or inputting incorrect data to obtain information or to influence the result of collaborative computations. From the transfer point of view, we plan to apply transparent, privacy-friendly machine learning in significant application domains, such as medicine, surveying, demand prediction and recommendation. In this context, we are interested to understand the appreciation of humans of transparency, verifiability, fairness, privacy-preserving and other trust-increasing aspects of our technologies.

4 Application domains

Our application domains cover health, mobility, social sciences and voice technologies.

Health Privacy is of major importance in the health domain. We contribute to develop methods to give access to the use of data in a private way rather than to the data itself centralized in vulnerable single locations. As an example, we are working with hospitals to develop the means of multicentric studies with privacy guarantees. A second example is personalized medicine where personal devices collect private and highly sensitive data. Potential applications of our research allow to keep data on device and to privately compute statistics.

Social sciences Our NLP research activities are rooted in linguistics, but learning unbiased representations of texts for instance or simply identifying unfair representations also have impacts in political sciences and history.

Music information retrieval By using analogies between language and music (symbolic notation) we tackle music information retrieval tasks such as style classification and structure detection.

Voice technologies We develop methods for privacy in speech that can be embedded in software suites dedicated to voice-based interaction systems.

5 Social and environmental responsibility

5.1 Footprint of research activities

Some of our research activities are energy intensive and we will work to reduce this carbon footprint in the future. Parts of the research projects FedMalin (see Section 10.3.2) and FLUTE are dedicated to this objective for the Federated Learning setting. In a collaboration with the Spirals team, we have extended the DecLearn API with features that are dedicated to energy consumption measurement with the PowerAPI

library. We are working on designing active strategies to select and schedule client participation in Federated learning, based on their energy consumption. The objective is to better handle the trade-off between energy consumption and accuracy in settings where the energy budget is limited.

5.2 Impact of research results

The main research topics of the team contribute to improve transparency, fairness and privacy in machine learning and reduce bias in natural language processing.

6 Highlights of the year

- Nicolas Papernot now holds INRIA international chair visiting Premedical, Magnet and Privatics.
- Two projects have been accepted for a pluridisciplinary research: CDP Loop and Prime Next-Gen. They will support the application of our research to the health domain.

7 Latest software developments, platforms, open data

7.1 Latest software developments

7.1.1 CoRTeX

Name: Python library for noun phrase COreference Resolution in natural language TEXTs

Keyword: Natural language processing

Functional Description: CoRTeX is a LGPL-licensed Python library for Noun Phrase coreference resolution in natural language texts. This library contains implementations of various state-of-the-art coreference resolution algorithms, including those developed in our research. In addition, it provides a set of APIs and utilities for text pre-processing, reading the CONLL2012 and CONLLU annotation formats, and performing evaluation, notably based on the main evaluation metrics (MUC, B-CUBED, and CEAF). As such, CoRTeX provides benchmarks for researchers working on coreference resolution, but it is also of interest for developers who want to integrate a coreference resolution within a larger platform. It currently supports use of the English or French language.

Contact: Pascal Denis

Participant: Pascal Denis

7.1.2 Mangoes

Name: MAGnet liNGuistic wOrd vEctorS

Functional Description: Mangoes is a toolbox for constructing and evaluating static and contextual token vector representations (aka embeddings). The main functionalities are:

- Contextual embeddings: Access a large collection of pretrained transformer-based language models, Pre-train a BERT language model on a corpus, Fine-tune a BERT language model for a number of extrinsic tasks, Extract features/predictions from pretrained language models.
 - Static embeddings: Process textual data and compute vocabularies and co-occurrence matrices. Input data should be raw text or annotated text, Compute static word embeddings with different state-of-the-art unsupervised methods, Propose statistical and intrinsic evaluation methods, as well as some visualization tools, Generate context dependent embeddings from a pretrained language model.
- Future releases will include methods for injecting lexical and semantic knowledge into token and multi-model embeddings, and interfaces into common external knowledge resources.

URL: <https://gitlab.inria.fr/magnet/mangoes>

Contact: Nathalie Vauquier

7.1.3 metric-learn

Keywords: Machine learning, Python, Metric learning

Functional Description: Distance metrics are widely used in the machine learning literature. Traditionally, practitioners would choose a standard distance metric (Euclidean, City-Block, Cosine, etc.) using a priori knowledge of the domain. Distance metric learning (or simply, metric learning) is the sub-field of machine learning dedicated to automatically constructing optimal distance metrics.

This package contains efficient Python implementations of several popular metric learning algorithms.

URL: <https://github.com/scikit-learn-contrib/metric-learn>

Contact: Aurélien Bellet

Partner: Parietal

7.1.4 MyLocalInfo

Keywords: Privacy, Machine learning, Statistics

Functional Description: Decentralized algorithms for machine learning and inference tasks which (1) perform as much computation as possible locally and (2) ensure privacy and security by avoiding that personal data leaves devices.

Contact: Nathalie Vauquier

7.1.5 declearn

Keyword: Federated learning

Scientific Description: declearn is a python package providing with a framework to perform federated learning, i.e. to train machine learning models by distributing computations across a set of data owners that, consequently, only have to share aggregated information (rather than individual data samples) with an orchestrating server (and, by extension, with each other).

The aim of declearn is to provide both real-world end-users and algorithm researchers with a modular and extensible framework that:

- (1) builds on abstractions general enough to write backbone algorithmic code agnostic to the actual computation framework, statistical model details or network communications setup
- (2) designs modular and combinable objects, so that algorithmic features, and more generally any specific implementation of a component (the model, network protocol, client or server optimizer...) may easily be plugged into the main federated learning process - enabling users to experiment with configurations that intersect unitary features
- (3) provides with functioning tools that may be used out-of-the-box to set up federated learning tasks using some popular computation frameworks (scikit-learn, tensorflow, pytorch...) and federated learning algorithms (FedAvg, Scaffold, FedYogi...)
- (4) provides with tools that enable extending the support of existing tools and APIs to custom functions and classes without having to hack into the source code, merely adding new features (tensor libraries, model classes, optimization plug-ins, orchestration algorithms, communication protocols...) to the party.

Parts of the declearn code (Optimizers,...) are included in the FedBioMed software.

At the moment, declearn has been focused on so-called "centralized" federated learning that implies a central server orchestrating computations, but it might become more oriented towards decentralized processes in the future, that remove the use of a central agent.

Functional Description: This library provides the two main components to perform federated learning:

- (1) the client, to be run by each participant, performs the learning on local data et releases only the result of the computation
- (2) the server orchestrates the process and aggregates the local models in a global model

News of the Year: Two major releases with key new functionalities including algorithms for group fairness and the ability to use secure aggregation.

URL: <https://gitlab.inria.fr/magnet/declearn/declearn2>

Contact: Aurélien Bellet

Participants: Paul Andrey, Aurélien Bellet, Nathan Bigaud, Marc Tommasi, Nathalie Vauquier

Partner: CHRU Lille

7.1.6 fairgrad

Name: FairGrad: Fairness Aware Gradient Descent

Keywords: Fairness, Fair and ethical machine learning, Machine learning, Classification

Functional Description: FairGrad is an easy to use general purpose approach in Machine Learning to enforce fairness in gradient descent based methods

URL: <https://github.com/saist1993/fairgrad>

Contact: Michael Perrot

7.1.7 tasksource

Name: tasksource

Keyword: Natural language processing

Functional Description: tasksource streamlines interchangeable datasets usage to scale evaluation or multi-task learning. All implemented preprocessings are in tasks.py or tasks.md. A preprocessing is a function that accepts a dataset and returns the standardized dataset. Preprocessing code is concise and human-readable.

URL: <https://github.com/sileod/tasksource>

Publication: [hal-04099649v1](https://hal.archives-ouvertes.fr/hal-04099649v1)

Contact: Damien Sileo

7.1.8 Voice Transformer 2

Keywords: Speech, Privacy

Scientific Description: The implemented method is inspired from the speaker anonymisation method proposed in [Fan+19], which performs voice conversion based on x-vectors [Sny+18], a fixed-length representation of speech signals that form the basis of state-of-the-art speaker verification systems. We have brought several improvements to this method such as pitch transformation, and new design choices for x-vector selection

[Fan+19] F. Fang, X. Wang, J. Yamagishi, I. Echizen, M. Todisco, N. Evans, and J.F. Bonastre. “Speaker Anonymization Using x-vector and Neural Waveform Models”. In: Proceedings of the 10th ISCA Speech Synthesis Workshop. 2019, pp. 155–160. [Sny+18] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur. “X-vectors: Robust DNN embeddings for speaker recognition”. In: Proceedings of ICASSP 2018 - 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2018, pp. 5329–5333.

Functional Description: Voice Transformer increases the privacy of users of voice interfaces by converting their voice into another person’s voice without modifying the spoken message. It ensures that any information extracted from the transformed voice can hardly be traced back to the original speaker, as validated through state-of-the-art biometric protocols, and it preserves the phonetic information required for human labelling and training of speech-to-text models.

Contact: Nathalie Vauquier

Participants: Brij Mohan Lal Srivastava, Nathalie Vauquier, Emmanuel Vincent, Marc Tommasi

7.2 Open data

Participants: Damien Sileo , Valentin Lacombe , Valentin Quesnel.

Reasoning Core (rc0, rc1)

Contributors: Damien Sileo, Valentin Lacombe, Valentin Quesnel

Description: Reasoning Core is a scalable environment for Reinforcement Learning with Verifiable Rewards (RLVR) designed to advance foundational symbolic reasoning in LLMs. It procedurally generates problems across core formal domains, including PDDL planning, first-order logic, context-free grammar parsing, causal reasoning, and system equation solving. This release includes the generated datasets rc0 and rc1, designed respectively for pre-training/mid-training (SFT) and post-training (RL).

Dataset PID (DOI,...): arXiv:2509.18083

Project link: <https://huggingface.co/datasets/reasoning-core/rc0>, <https://huggingface.co/datasets/reasoning-core/rc1>

Publications: V. Lacombe, V. Quesnel, D. Sileo. Reasoning Core: A Scalable RL Environment for LLM Symbolic Reasoning. arXiv preprint arXiv:2509.18083, 2025. [48]

Contact: Damien Sileo

Release contributions: Datasets and Code environment for LLM symbolic reasoning

8 New results

8.1 Natural Language Processing

Participants: Damien Sileo , Bastien Liétard , Valentin Lacombe , Angèle Brunellière.

Reasoning Core: A Scalable RL Environment for LLM Symbolic Reasoning [48]

We introduce Reasoning Core, a new scalable environment for Reinforcement Learning with Verifiable Rewards (RLVR), designed to advance foundational symbolic reasoning in Large Language Models (LLMs). Unlike existing benchmarks that focus on games or isolated puzzles, Reasoning Core procedurally generates problems across core formal domains, including PDDL planning, first-order logic, context-free grammar parsing, causal reasoning, and system equation solving. The environment is built on key design principles of high-generality problem distributions, verification via external tools, and continuous difficulty control, which together provide a virtually infinite supply of novel training instances. Initial zero-shot evaluations with frontier LLMs confirm the difficulty of Reasoning Core’s tasks, positioning it as a promising resource to improve the reasoning capabilities of future models.

Bridging the Data Provenance Gap Across Text, Speech and Video [33]

Progress in AI is driven largely by the scale and quality of training data. Despite this, there is a deficit of empirical analysis examining the attributes of well-established datasets beyond text. In this work we conduct the largest and first-of-its-kind longitudinal audit across modalities—popular text, speech, and video datasets—from their detailed sourcing trends and use restrictions to their geographical and linguistic representation. Our manual analysis covers nearly 4000 public datasets between 1990-2024, spanning 608 languages, 798 sources, 659 organizations, and 67 countries. We find that multimodal machine learning applications have overwhelmingly turned to web-crawled, synthetic, and social media platforms, such as YouTube, for their training sets, eclipsing all other sources since 2019. Secondly, tracing the chain of dataset derivations we find that while less than 33% of datasets are restrictively licensed, over 80% of the source content in widely-used text, speech, and video datasets, carry non-commercial restrictions. Finally, counter to the rising number of languages and geographies represented in public AI training datasets, our audit demonstrates measures of relative geographical and multilingual representation have failed to significantly improve their coverage since 2013. We believe the breadth of our audit enables us to empirically examine trends in data sourcing, restrictions, and Western-centricity at an ecosystem-level, and that visibility into these questions are essential to progress in responsible AI. As a contribution to ongoing improvements in dataset transparency and responsible use, we release our entire multimodal audit, allowing practitioners to trace data provenance across text, speech, and video.

Humanity’s Last Exam [23]

Benchmarks are important tools for tracking the rapid advancements in large language model (LLM) capabilities. However, benchmarks are not keeping pace in difficulty: LLMs now achieve over 90% accuracy on popular benchmarks like MMLU, limiting informed measurement of state-of-the-art LLM capabilities. In response, we introduce Humanity’s Last Exam (HLE), a multi-modal benchmark at the frontier of human knowledge, designed to be the final closed-ended academic benchmark of its kind with broad subject coverage. HLE consists of 3,000 questions across dozens of subjects, including mathematics, humanities, and the natural sciences. HLE is developed globally by subject-matter experts and consists of multiple-choice and short-answer questions suitable for automated grading. Each question has a known solution that is unambiguous and easily verifiable, but cannot be quickly answered via internet retrieval. State-of-the-art LLMs demonstrate low accuracy and calibration on HLE, highlighting a significant gap between current LLM capabilities and the expert human frontier on closed-ended academic questions. To inform research and policymaking upon a clear understanding of model capabilities, we publicly release [HLE](#).

Saturation-Driven Dataset Generation for LLM Mathematical Reasoning in the TPTP Ecosystem [50]

The scarcity of high-quality, logically sound data is a critical bottleneck for advancing the mathematical reasoning of Large Language Models (LLMs). Our work confronts this challenge by turning decades of automated theorem proving research into a scalable data engine. Rather than relying on error-prone LLMs or complex proof-assistant syntax like Lean and Isabelle, our framework leverages E-prover’s saturation capabilities on the vast TPTP axiom library to derive a massive, guaranteed-valid corpus of theorems. Our pipeline is principled and simple: saturate axioms, filter for “interesting” theorems, and generate tasks. With no LLMs in the loop, we eliminate factual errors by construction. This purely symbolic data is then transformed into three difficulty-controlled challenges: entailment verification, premise selection, and proof reconstruction. Our zero-shot experiments on frontier models reveal a clear weakness: performance collapses on tasks requiring deep, structural reasoning. Our framework provides both the diagnostic tool to measure this gap and a scalable source of symbolic training data to address it. We make the code and data [publicly available](#).

Logic Haystacks: Probing LLMs Long-Context Logical Reasoning (Without Easily Identifiable Unrelated Padding) [53]

Large language models demonstrate promising long context processing capabilities, with recent models touting context windows close to one million tokens. However, the evaluations supporting these claims often involve simple retrieval tasks or synthetic tasks padded with irrelevant text, which the models may

easily detect and discard. In this work, we generate lengthy simplified English text with first-order logic representations spanning up to 2048 clauses (around 25k GPT-4 tokens). We formulate an evaluation task with evidence retrieval for contradiction detection. The long, homogeneous text is filled with distractors that are both hard to distinguish from relevant evidences and provably not interfering with them. Our evaluation of evidence retrieval shows that the effective context window is much smaller with realistic distractors, already crumbling at 128 clauses.

Generating Explanations in Medical Question-Answering by Expectation Maximization Inference over Evidence [24]

Medical Question Answering (medical QA) systems play an essential role in assisting healthcare workers in finding answers to their questions. However, it is not sufficient to merely provide answers by medical QA systems because users might want explanations, that is, more analytic statements in natural language that describe the elements and context that support the answer. To do so, we propose a novel approach for generating natural language explanations for answers predicted by medical QA systems. As high-quality medical explanations require additional medical knowledge, so that our system extracts knowledge from medical textbooks to enhance the quality of explanations during the explanation generation process. Concretely, we designed an Expectation-Maximization approach that makes inferences about the evidence found in these texts, offering an efficient way to focus attention on lengthy evidence passages. Experimental results, conducted on two datasets MQAE-diag and MQAE, demonstrate the effectiveness of our framework for reasoning with textual evidence. Our approach outperforms state-of-the-art models, achieving a significant improvement of 6.13 and 5.47 percentage points on the Rouge-L score; 6.49 and 5.28 percentage points on the Bleu-4 score on the MQAE-diag and MQAE datasets.

Neural evidence for perceiving a vowel merger after a social interaction within a native language [17]

Although previous research has shown that speakers adapt on the words they use, it remains unclear whether speakers adapt their phonological representations, leading them to perceive new phonemic contrasts following a social interaction. This event-related potential (ERP) study investigates whether the neuronal responses to the perception of the /e/-/ɛ/ vowel merger in Northern French speakers show evidence for discriminating /e/ and /ɛ/ phonemes after interacting with a speaker who produced this contrast. Northern French participants engaged in an interactive map task and we measured their ERP responses elicited after the presentation of a last syllable which was either phonemically identical to or different from preceding syllables. There was no evidence for discrimination between /e/ and /ɛ/ phonemes before the social interaction, while mismatch negativity (MMN) and late responses revealed /e/-/ɛ/ discrimination after the social interaction. The findings suggest rapid neuronal adaptations of phonemic representations thanks to the social interaction.

How does the creation of new semantic relationships during dialogue impact long-term semantic representations after dialogue? [19]

Dialogue is an ideal setting for changing linguistic representations thanks to the repeated use of new words and meanings. Two experiments were conducted to examine the extent to which new semantic relationships created during dialogue may change preexisting representations in long-term semantic memory after a dialogue. For this purpose, we developed an interactive agreement referential task to create new semantic relationships in dialogue between two words by associating them to a single picture. One day after the dialogue phase, participants performed either a lexical decision task associated to a semantic priming paradigm (Experiment 1), or a semantic relatedness judgment task (Experiment 2). In both tasks, the participants' performance was collected during the processing of pairs of words referring either to the same picture or to different pictures during the dialogue phase in order to assess changes in long-term semantic representations after the dialogue phase. No significant effect of relatedness due to the creation of new semantic relationships during dialogue was found in the lexical decision task. However, when the participants' attention was focused on semantic relationships during the semantic relatedness judgment task, which required participants to perform an explicit judgment, newly related words were rated as more related semantically. The two experiments bear important implications for understanding on the links between dialogue and the updating of long-term semantic representations.

Indirect Reply Processing in Multilingual Conversations when Inferring Speaker Meaning: an ERP study [37, 56]

For a successful communication, interlocutors need to interpret an intended meaning beyond what is explicitly stated. Accordingly, a pragmatic inference is often required, as is the case for indirect replies. For instance, if someone replies to "What is it like giving a presentation?" with "Giving a good presentation is complicated", the reply is direct. However, when responding to "Did you like my presentation?", the same reply becomes indirect, implying the presentation was not well-received. An intriguing question is to know whether such pragmatic inferences, required for indirect reply processing, are influenced by the cognitive demands of processing a second language (L2). To explore this question, our experiment aimed to characterize the neural signature of indirect reply processing in L1 and L2 dialogues. We hypothesized that the additional cognitive cost of L2 processing would hinder pragmatic inferencing. To test this, we measured the event-related potentials (ERPs) of 40 French-speaking students listening to 144 dialogues in French (L1) and English (L2), in a within-participants design. Each dialogue, preceded by a written context, ended with a direct or indirect reply. The ERPs were time-locked on the final word of the reply. A visual inspection of the results revealed that L1 dialogues elicited an early ERP effect, indicating faster semantic processing compared to L2 dialogues. Initially, the response was more negative for direct replies; however, after 400 ms, this negative response became larger for indirect replies, reflecting an increased cognitive effort required for pragmatic inferences. In L2 dialogues, indirect replies elicited a persistent negative response after 400 ms, suggesting greater difficulty when inferring the implied meaning in L2 dialogues. These findings highlight the additional cognitive demands of processing indirect replies, particularly in L2 over late-processing stages. The current study therefore contributes to our understanding of pragmatic processing and its neural underpinnings in multilingual communication.

Interacting with someone shapes prediction in spoken-language comprehension [54]

Background: While listening to a spoken message, predicting upcoming information from a previous sentential context ensures a successful comprehension [1,2]. However, prediction in spoken-language comprehension is not always found [3,4]. In order to explain this flexibility, both prediction-by-association and prediction-by-production have been proposed [2]. While prediction-by-association would be an automatic process related to the spreading activation of conceptual features from the sentence representation, prediction-by-production, leading to the preactivation of words predicted from the sentence representation and their properties thanks to the production system, would be optional. In this study, we investigated the flexibility of prediction-by-association and prediction-by-production through a social interaction. This investigation is motivated by two key components of social interactions: mutual comprehension at the conceptual level and interplay between comprehension and production systems.
Method: In the Visual World Paradigm, thirty-two native French speakers listened to forty-eight highly and weakly constraining sentences which were associated with a visual scene containing three distractor objects and one of four critical objects (Target, Semantic Competitor, Phonological Competitor, Unrelated). Each sentence was presented only once before or after a social interaction, which was not related to the content of the sentences (see Figure 1A). The Target object referred to the word predicted from a sentence whose effect could reflect the two types of predictions. The Semantic and Phonological Competitor objects referred respectively to a word sharing either semantic category or phonological onset overlap with the predicted word to explore prediction-by-production. Used as a control condition, the Unrelated object referred to a word with neither phonological onset overlap nor semantic relationships with the predicted word. Fixations on the different objects were recorded during sentence comprehension. After hearing the sentence, participants had to determine whether one of the objects was mentioned in the sentence. During the social interaction, participants had to find the correct position for five Tangram pictures into a grid of ten pictures under time pressure by actively collaborating with their partner and describing shapes via an audioconference device.
Results: Cluster-based permutation analyses were performed on fixation proportion differences between related and unrelated objects. For highly constraining sentences, fixation proportions between Target and Unrelated objects differed in a time window between -320 and 980 ms after word onset before the interaction ($p=.001$, see Figure 1B). This difference emerged 160 ms earlier after the interaction in a time window between -480 ms and 980 after word onset. Fixation proportions between Semantic Competitor and Unrelated objects differed in a time window between -380 and -80 ms after word onset before the interaction ($p=.004$, see Figure 1B). No other

significant effects were found for highly or weakly constraining sentences. Discussion: Consistent with optional prediction-by-production [2], a predictive effect due to Semantic Competitor objects only occurred before the social interaction. In contrast, the predictive effect of Target objects was speeded-up after the social interaction. This pilot study suggests that social interaction may shift prediction toward finer based-concept processing. Findings will be discussed in line with models of social interaction and language comprehension and further studies are needed to confirm the role of interactions.

Neural signature of indirect reply processing while listening to foreign accented dialogues [55]

The interpretation of the sentence ‘Giving a good presentation is complicated’ will differ whether it is a reply to (a) or (b): (a) What is it like giving a presentation? (b) Did you like my presentation? The neurocognitive mechanisms underlying the pragmatic inference of indirect replies (e.g. in reply to (b)), have mainly been studied when dialogues occur in first language contexts. However, we hypothesize that foreign-accented speech may affect such inferences given the cognitive cost it generates (due to linguistic disfluency), as well as native listeners’ limited expectation of the foreign speaker’s linguistic abilities. To test this hypothesis, our ongoing experiment aims to characterize the neural signature of indirect reply processing in foreign-accented dialogues. Accordingly, we measure the event-related potentials (ERPs) of 40 French-speaking students listening to native and foreign-accented dialogues. Each dialogue is preceded by a written context establishing the communicative situation. A total of 144 dialogues are presented, ending in either a direct or indirect reply (e.g. in reply to (a) or (b), respectively). For one-third of the dialogues, participants answer yes-no comprehension questions. Based on the few previous ERP studies of indirect replies and foreign-accented pragmatics, we expect to find ERPs time-locked to the final word of the reply that reflect additional cognitive effort when processing foreign-accented indirect replies. Compared to native-accented direct replies, we anticipate a shallower processing. Particularly, we should observe a smaller and delayed N400, indicating a hindered semantic integration of the reply, and a larger P600, reflecting the additional pragmatic processing. Moreover, we will examine explanatory hypotheses of how individual cognitive capacities, such as working memory and non-verbal reasoning, may modulate pragmatic processing in multilingual communication, as previously observed in native-accented contexts. Our results will be discussed in line with the neurocognitive models of language comprehension and pragmatics.

Faces and voices in dialogue: How partner-specific cues contribute to conversational memory [38]

Previous research suggests that information mentioned during dialogue is frequently encoded in association with the current partner. This raises the question of which partner-specific cues might contribute to the subsequent retrieval of information from memory. Following a joint communication task, individuals were tested on recognition memory for referent labels in a context cued by their partner’s face and/or voice. We examine whether partner-specific visual and auditory cues can facilitate access to information encoded during conversation.

8.2 Privacy and NLP

Participants: Marc Tommasi , Damien Sileo , Gabriel Loiseau.

Tau-Eval: A Unified Evaluation Framework for Useful and Private Text Anonymization [32]

Text anonymization is the process of removing or obfuscating information from textual data to protect the privacy of individuals. This process inherently involves a complex trade-off between privacy protection and information preservation, where stringent anonymization methods can significantly impact the text’s utility for downstream applications. Evaluating the effectiveness of text anonymization proves challenging from both privacy and utility perspectives, as there is no universal benchmark that can comprehensively assess anonymization techniques across diverse, and sometimes contradictory contexts. We present Tau-Eval, an open-source framework for benchmarking text anonymization methods through the lens of privacy and utility task sensitivity. A Python library, code, documentation and tutorials are publicly available.

TAROT: Task-Oriented Authorship Obfuscation Using Policy Optimization Methods [31]

Authorship obfuscation aims to disguise the identity of an author within a text by altering the writing style, vocabulary, syntax, and other linguistic features associated with the text author. This alteration needs to balance privacy and utility. While strong obfuscation techniques can effectively hide the author's identity, they often degrade the quality and usefulness of the text for its intended purpose. Conversely, maintaining high utility tends to provide insufficient privacy, making it easier for an adversary to de-anonymize the author. Thus, achieving an optimal trade-off between these two conflicting objectives is crucial. In this paper, we propose TAROT: Task-Oriented Authorship Obfuscation Using Policy Optimization, a new unsupervised authorship obfuscation method whose goal is to optimize the privacy-utility trade-off by regenerating the entire text considering its downstream utility. Our approach leverages policy optimization as a fine-tuning paradigm over small language models in order to rewrite texts by preserving author identity and downstream task utility. We show that our approach largely reduce the accuracy of attackers while preserving utility. We make our code and models publicly available.

8.3 Music and NLP

Participants: Mikaela Keller , Dinh Viet-Toan Le.

Natural Language Processing Methods for Symbolic Music Generation and Information Retrieval: a Survey [21]

Several adaptations of Transformers models have been developed in various domains since its breakthrough in Natural Language Processing (NLP). This trend has spread into the field of Music Information Retrieval (MIR), including studies processing music data. However, the practice of leveraging NLP tools for symbolic music data is not novel in MIR. Music has been frequently compared to language, as they share several similarities, including sequential representations of text and music. These analogies are also reflected through similar tasks in MIR and NLP. This survey reviews NLP methods applied to symbolic music generation and information retrieval studies following two axes. We first propose an overview of representations of symbolic music adapted from natural language sequential representations. Such representations are designed by considering the specificities of symbolic music. These representations are then processed by models. Such models, possibly originally developed for text and adapted for symbolic music, are trained on various tasks. We describe these models, in particular deep learning models, through different prisms, highlighting music-specialized mechanisms. We finally present a discussion surrounding the effective use of NLP tools for symbolic music data. This includes technical issues regarding NLP methods and fundamental differences between text and music, which may open several doors for further research into more effectively adapting NLP tools to symbolic MIR.

METEOR: Melody-aware Texture-controllable Symbolic Orchestral Music Generation via Transformer VAE [29]

Re-orchestration is the process of adapting a music piece for a different set of instruments. By altering the original instrumentation, the orchestrator often modifies the musical texture while preserving a recognizable melodic line and ensures that each part is playable within the technical and expressive capabilities of the chosen instruments. In this work, we propose METEOR, a model for generating Melody-aware Texture-controllable re-Orchestration with a Transformer-based variational auto-encoder (VAE). This model performs symbolic instrumental and textural music style transfers with a focus on melodic fidelity and controllability. We allow bar- and track-level controllability of the accompaniment with various textural attributes while keeping a homophonic texture. With both subjective and objective evaluations, we show that our model outperforms style transfer models on a re-orchestration task in terms of generation quality and controllability. Moreover, it can be adapted for a lead sheet orchestration task as a zero-shot learning model, achieving performance comparable to a model specifically trained for this task.

Evaluating Interval-based Tokenization for Pitch Representation in Symbolic Music Analysis [57]

Symbolic music analysis tasks are often performed by models originally developed for Natural Language Processing, such as Transformers. Such models require the input data to be represented as sequences, which is achieved through a process of tokenization. Tokenization strategies for symbolic music often rely on absolute MIDI values to represent pitch information. However, music research largely promotes the benefit of higher-level representations such as melodic contour and harmonic relations for which pitch intervals turn out to be more expressive than absolute pitches. In this work, we introduce a general framework for building interval-based tokenizations. By evaluating these tokenizations on three music analysis tasks, we show that such interval-based tokenizations improve model performances and facilitate their explainability.

Modeling Symbolic Music with Natural Language Processing Approaches [41]

Music is often described as a language because of its similarities to natural language. These include their respective representations through symbolic music notation and textual form. Therefore, the field of Music Information Retrieval (MIR) has often borrowed several tools from the Natural Language Processing (NLP) field to adapt them to process symbolic music data. In particular, this phenomenon has been increasingly popular with the breakthrough of Transformer models in the NLP field. This thesis first provides a structured overview of adaptations of NLP methods developed in the MIR field for symbolic music processing. They are presented along three axes, each addressing the use of diverse representations of symbolic music at different levels. Symbolic music represented as sequential data has led to the development of several tokenization strategies, which we propose to organize within a unified taxonomy. These representations are subsequently processed through models, such as recurrent or attention-based architectures initially developed for text data, giving rise to multiple adaptations for symbolic music processing. Finally, these abstract representations are used to perform tasks, where both parallels and distinctive characteristics emerge between MIR and NLP. These aspects then structure the three technical contributions of this thesis. First, we study the expressiveness of sequential representations of music through the development of interval-based tokenization strategies, and the analysis of a subword tokenization strategy, Byte-Pair Encoding, applied to symbolic music tokens. We then propose a framework for model explainability which leads to the analysis of the attention mechanism of a Transformer-based model trained for functional harmony analysis. Finally, we develop a model adapted from NLP tools for a task of re-orchestration, framed as a case of multi-track music generation. Ultimately, this thesis defends that NLP methods first remains a toolbox from which MIR studies can take some tools from. Beyond the analogies between music and natural language, the main motivation guiding a MIR study should be musical questions.

8.4 Speech and Privacy

Participants: Marc Tommasi.

Analysis of Speech Temporal Dynamics in the Context of Speaker Verification and Voice Anonymization [35]

In this paper, we investigate the impact of speech temporal dynamics in application to automatic speaker verification and speaker voice anonymization tasks. We propose several metrics to perform automatic speaker verification based only on phoneme durations. Experimental results demonstrate that phoneme durations leak some speaker information and can reveal speaker identity from both original and anonymized speech. Thus, this work emphasizes the importance of taking into account the speaker's speech rate and, more importantly, the speaker's phonetic duration characteristics, as well as the need to modify them in order to develop anonymization systems with strong privacy protection capacity.

Exploiting Context-dependent Duration Features for Voice Anonymization Attack Systems [36]

The temporal dynamics of speech, encompassing variations in rhythm, intonation, and speaking rate, contain important and unique information about speaker identity. This paper proposes a new method for representing

speaker characteristics by extracting context-dependent duration embeddings from speech temporal dynamics. We develop novel attack models using these representations and analyze the potential vulnerabilities in speaker verification and voice anonymization systems. The experimental results show that the developed attack models provide a significant improvement in speaker verification performance for both original and anonymized data in comparison with simpler representations of speech temporal dynamics reported in the literature.

8.5 Security and Privacy

Participants: Marc Tommasi , Jan Ramon , Antoine Barczewski , Marc Damie , Clément Pierquin , Paul Andrey , Michaël Perrot , Jean Dufranche.

Generalization under Byzantine and Poisoning Attacks: Tight Stability Bounds in Robust Distributed Learning [42]

Robust distributed learning algorithms aim to maintain good performance in distributed and federated settings, even in the presence of misbehaving workers. Two primary threat models have been studied: Byzantine attacks, where misbehaving workers can send arbitrarily corrupted updates, and data poisoning attacks, where misbehavior is limited to manipulation of local training data. While prior work has shown comparable optimization error under both threat models, a fundamental question remains open: How do these threat models impact generalization? Empirical evidence suggests a gap between the two threat models, yet it remains unclear whether it is fundamental or merely an artifact of suboptimal attacks. In this work, we present the first theoretical investigation into this problem, formally showing that Byzantine attacks are intrinsically more harmful to generalization than data poisoning. Specifically, we prove that: (i) under data poisoning, the uniform algorithmic stability of a robust distributed learning algorithm, with optimal optimization error, degrades by an additive factor of $\Theta(\frac{f}{n-f})$, with f the number of misbehaving workers out of n ; and (ii) In contrast, under Byzantine attacks, the degradation is in $O(\sqrt{\frac{f}{n-2f}})$. This difference in stability leads to a generalization error gap that is especially significant as f approaches its maximum value $\frac{n}{2}$.

Privacy-Preserving Computations on Sparse Data [40]

Data breaches and privacy violations have raised global concerns about the protection of personal information. To address these concerns, cryptographic protocols known as Multi-Party Computations (MPC) have been developed to enable multiple parties to jointly compute on their private inputs without revealing them. These protocols have found applications notably in tax fraud detection, healthcare, and machine learning. However, existing MPC protocols remain insufficient for many real-world scenarios, particularly when dealing with high-dimensional or structured data. This thesis focuses on sparse data; datasets containing mostly zero values, which naturally arise in applications such as recommender systems and healthcare. While plaintext algorithms for sparse data are well established, few cryptographic protocols are optimized for this setting, limiting the practicality of MPC in domains where sparsity is the norm. The core contributions of this thesis introduce new cryptographic protocols tailored to sparse computations. We propose MPC protocols for secure sparse matrix multiplication, enabling performance gains that make previously impractical applications feasible. We also design new Function Secret Sharing (FSS) schemes able to efficiently aggregate sparse data. Beyond these protocols, the thesis makes several orthogonal contributions that question key assumptions and practical aspects of privacy-enhancing technologies. We study the role of leakage in sparse computations by analyzing access-pattern leakage in searchable encryption, providing new attacks and statistical insights into its risks. We present Fedivertex, a new graph dataset based on decentralized social media, to benchmark decentralized machine learning. Finally, we evaluate the energy consumption of several privacy-enhancing technologies. Taken together, these contributions both advance the design of cryptographic protocols for sparse data and provide a broader perspective on the challenges of deploying privacy-preserving computations in practice.

Noisy Function Secret Sharing and its applications to Differentially Private computations [45]

Function Secret Sharing (FSS) schemes enable to share secret functions between multiple parties, with notable applications in anonymous communication and privacy-preserving machine learning. While two-party schemes offer logarithmic key sizes, multi-party schemes remain less practical due to significantly larger keys. Although several approaches have been proposed to improve multi-party schemes, a significant efficiency gap remains between the two-party and multi-party settings.

Our work introduces noisy FSS: a relaxation of FSS preserving the standard privacy guarantees but relaxing the correctness definition by allowing a small amount of noise in the output. We formally define noisy FSS and show how the noise introduced by the scheme can be leveraged to provide differential private outputs in statistics applications.

To demonstrate the benefits of this relaxation, we adapt a scheme proposed by Corrigan-Gibbs et al. (S&P'15). While their scheme provides the smallest key sizes among multi-party schemes, they do not support some applications notably in statistics due to their non-linear share decoding. On the contrary, recent works such as Goel et al. (CRYPTO'25) have larger keys, but support all FSS applications. Our noisy adapted scheme offers the best of both worlds by matching the best key sizes, while providing the properties necessary to statistics applications.

How to Securely Shuffle? A survey about Secure Shufflers for privacy-preserving computations [43]

Ishai et al. (FOCS'06) introduced secure shuffling as an efficient building block for private data aggregation. Recently, the field of differential privacy has revived interest in secure shufflers by highlighting the privacy amplification they can provide in various computations. Although several works argue for the utility of secure shufflers, they often treat them as black boxes; overlooking the practical vulnerabilities and performance trade-offs of existing implementations. This leaves a central question open: what makes a good secure shuffler? This survey addresses that question by identifying, categorizing, and comparing 26 secure protocols that realize the necessary shuffling functionality. To enable a meaningful comparison, we adapt and unify existing security definitions into a consistent set of properties. We also present an overview of privacy-preserving technologies that rely on secure shufflers, offer practical guidelines for selecting appropriate protocols, and outline promising directions for future work.

Revisiting the Attacker's Knowledge in Inference Attacks Against Searchable Symmetric Encryption [27]

Encrypted search schemes have been proposed to address growing privacy concerns. However, several leakage-abuse attacks have highlighted some security vulnerabilities. Recent attacks assumed an attacker's knowledge containing data "similar" to the indexed data. However, this vague assumption is barely discussed in literature: how likely is it for an attacker to obtain a "similar enough" data? Our paper provides novel statistical tools usable on any attack in this setting to analyze its sensitivity to data similarity. First, we introduce a mathematical model based on statistical estimators to analytically understand the attackers' knowledge and the notion of similarity. Second, we conceive statistical tools to model the influence of the similarity on the attack accuracy. We apply our tools on three existing attacks to answer questions such as: is similarity the only factor influencing accuracy of a given attack? Third, we show that the enforcement of a maximum index size can make the "similar-data" assumption harder to satisfy. In particular, we propose a statistical method to estimate an appropriate maximum size for a given attack and dataset. For the best known attack on the Enron dataset, a maximum index size of 200 guarantees (with high probability) the attack accuracy to be below 5%.

Secure Sparse Matrix Multiplications and their Applications to Privacy-Preserving Machine Learning [44]

To preserve privacy, multi-party computation (MPC) enables executing Machine Learning (ML) algorithms on secret-shared or encrypted data. However, existing MPC frameworks are not optimized for sparse data. This makes them unsuitable for ML applications involving sparse data, e.g., recommender systems or genomics. Even in plaintext, such applications involve high-dimensional sparse data, that cannot be processed without sparsity-related optimizations due to prohibitively large memory requirements. Since

matrix multiplication is central in ML algorithms, we propose MPC algorithms to multiply secret sparse matrices. On the one hand, our algorithms avoid the memory issues of the "dense" data representation of classic secure matrix multiplication algorithms. On the other hand, our algorithms can significantly reduce communication costs (some experiments show a factor 1000) for realistic problem sizes. We validate our algorithms in two ML applications in which existing protocols are impractical. An important question when developing MPC algorithms is what assumptions can be made. In our case, if the number of non-zeros in a row is a sensitive piece of information then a short runtime may reveal that the number of non-zeros is small. Existing approaches make relatively simple assumptions, e.g., that there is a universal upper bound to the number of non-zeros in a row. This often doesn't align with statistical reality, in a lot of sparse datasets the amount of data per instance satisfies a power law. We propose an approach which allows adopting a safe upper bound on the distribution of non-zeros in rows/columns of sparse matrices.

Dropout-Robust Mechanisms for Differentially Private and Fully Decentralized Mean Estimation [52]

Achieving differentially private computations in decentralized settings poses significant challenges, particularly regarding accuracy, communication cost, and robustness against information leakage. While cryptographic solutions offer promise, they often suffer from high communication overhead or require centralization in the presence of network failures. Conversely, existing fully decentralized approaches typically rely on relaxed adversarial models or pairwise noise cancellation, the latter suffering from substantial accuracy degradation if parties unexpectedly disconnect. In this work, we propose IncA, a new protocol for fully decentralized mean estimation, a widely used primitive in data-intensive processing. Our protocol, which enforces differential privacy, requires no central orchestration and employs low-variance correlated noise, achieved by incrementally injecting sensitive information into the computation. First, we theoretically demonstrate that, when no parties permanently disconnect, our protocol achieves accuracy comparable to that of a centralized setting—already an improvement over most existing decentralized differentially private techniques. Second, we empirically show that our use of low-variance correlated noise significantly mitigates the accuracy loss experienced by existing techniques in the presence of dropouts.

TAMIS: Tailored Membership Inference Attacks on Synthetic Data [25]

Membership Inference Attacks (MIA) enable to empirically assess the privacy of a machine learning algorithm. In this paper, we propose TAMIS, a novel MIA against differentially-private synthetic data generation methods that rely on graphical models. This attack builds upon MAMA-MIA, a recently-published state-of-the-art method. It lowers its computational cost and requires less attacker knowledge. Our attack is the product of a two-fold improvement. First, we recover the graphical model having generated a synthetic dataset by using solely that dataset, rather than shadow-modeling over an auxiliary one. This proves less costly and more performant. Second, we introduce a more mathematically-grounded attack score, that provides a natural threshold for binary predictions. In our experiments, TAMIS achieves better or similar performance as MAMA-MIA on replicas of the SNAKE challenge.

Privacy Amplification Through Synthetic Data: Insights from Linear Regression [34]

Synthetic data inherits the differential privacy guarantees of the model used to generate it. Additionally, synthetic data may benefit from privacy amplification when the generative model is kept hidden. While empirical studies suggest this phenomenon, a rigorous theoretical understanding is still lacking. In this paper, we investigate this question through the well-understood framework of linear regression. First, we establish negative results showing that if an adversary controls the seed of the generative model, a single synthetic data point can leak as much information as releasing the model itself. Conversely, we show that when synthetic data is generated from random inputs, releasing a limited number of synthetic data points amplifies privacy beyond the model's inherent guarantees. We believe our findings in linear regression can serve as a foundation for deriving more general bounds in the future.

Enhancing Differentially private machine learning: Optimizations for Repeated Query scenarios [39]

Deep neural networks and other machine learning models have experienced unprecedented growth in recent years. Alongside this enthusiasm, there has been an increasing and well-founded concern about the privacy

of the vast amounts of data required to train these models. The combination of these two factors has been a key driver of interest in privacy-preserving machine learning techniques. Differential Privacy has emerged as the gold standard for measuring privacy. This framework is now applied on a wide range of data-driven tasks, such as machine learning and collaborative analysis, where multiple stakeholders wish to query shared data without exposing their own. The main challenge in this domain lies in balancing privacy guarantees with the utility of the results. Indeed, privacy-preserving techniques often come at the cost of reduced utility. This thesis focuses on techniques to improve machine learning models and tools for analyzing them, while ensuring a satisfactory level of privacy for the underlying data. First, it introduces an innovative approach to privacy-preserving gradient descent methods by addressing the bias introduced by existing methods. By leveraging properties of gradient regularity rather than clipping the gradient, as it is commonly done in popular methods, our approach effectively reduces bias and the noise added to the gradient. We propose a new algorithm that surpasses the state of the art across various datasets. Second, the thesis explores techniques for computing privacy-preserving empirical cumulative distribution functions, even in cases where the data is distributed across multiple entities. This study proposes a novel method compatible with different security protocols, offering provable privacy guarantees and an analysis of computational costs. A range of applications are explored, and experimental results are presented to validate the utility of these methods. By analyzing optimization mechanisms and distribution functions, this thesis contributes to the development of more practical and efficient privacy-preserving machine learning and data analysis techniques.

Evaluating Membership Inference Attacks in Heterogeneous-Data Setups [28]

Among all privacy attacks against Machine Learning (ML), membership inference attacks (MIA) attracted the most attention. In these attacks, the attacker is given an ML model and a data point, and they must infer whether the data point was used for training. The attacker also has an auxiliary dataset to tune their inference algorithm. Attack papers commonly simulate setups in which the attacker's and the target's datasets are sampled from the same distribution. This setting is convenient to perform experiments, but it rarely holds in practice. ML literature commonly starts with similar simplifying assumptions (i.e., "i.i.d." datasets), and later generalizes the results to support heterogeneous data distributions. Similarly, our work makes a first step in the generalization of the MIA evaluation to heterogeneous data. First, we design a metric to measure the heterogeneity between any pair of tabular data distributions. This metric provides a continuous scale to analyze the phenomenon. Second, we compare two methodologies to simulate a data heterogeneity between the target and the attacker. These setups provide opposite performances: 90% attack accuracy vs. 50% (i.e., random guessing). Our results show that the MIA accuracy depends on the experimental setup; and even if research on MIA considers heterogeneous data setups, we have no standardized baseline of how to simulate it. The lack of such a baseline for MIA experiments poses a significant challenge to risk assessments in real-world machine learning scenarios.

Learning with Locally Private Examples by Inverse Weierstrass Private Stochastic Gradient Descent (in preparation)

Local Differential Privacy (LDP) has emerged as a prevailing framework for publishing data privately without relying on a central trusted authority. In this paper, we are interested in such a setting where examples are first privatized using the Gaussian and Randomized Response mechanisms and then publicly released. We first leverage tools from the Gaussian smoothing literature, namely the Weierstrass transform, to characterize the bias that standard risk minimization on such privatized data would induce. To mitigate this bias, we then propose the Inverse Weierstrass Private SGD algorithm (IWP-SGD), a variant of stochastic gradient descent that is guaranteed to recover, in expectation, the solution to the original, non-biased problem. This new method leverages the invertibility of the Weierstrass transform to build an unbiased estimator of the gradient with finite variance. It allows us to derive non-asymptotic convergence rates. Empirically, we validate it on binary problems on synthetic and real data.

8.6 Fairness

Participants: Michaël Perrot , Marc Tommasi , Shreya Venugopal.

Fair Text Classification via Transferable Representations [22]

Group fairness is a central research topic in text classification, where reaching fair treatment between sensitive groups (e.g., women and men) remains an open challenge. We propose an approach that extends the use of the Wasserstein Dependency Measure for learning unbiased neural text classifiers. Given the challenge of distinguishing fair from unfair information in a text encoder, we draw inspiration from adversarial training by inducing independence between representations learned for the target label and those for a sensitive attribute. We further show that Domain Adaptation can be efficiently leveraged to remove the need for access to the sensitive attributes in the dataset we cure. We provide both theoretical and empirical evidence that our approach is well-founded.

Preserving Fairness when Making Stochastic Predictions Deterministic (in preparation)

Deterministic decisions are desirable in many high-stake automated decision processes. The usual solution to achieve that when only stochastic classifiers are available is to use 0.5 thresholding. Unfortunately, this process may induce additional unfairness. In this paper, we propose the first method specifically tailored to make stochastic predictions deterministic while preserving the fairness level of the stochastic classifier, that is without introducing any additional bias. Leveraging ideas from the post-processing literature, we demonstrate that our method is theoretically sound. It comes with generalization guarantees for both fairness and accuracy. Furthermore, it inherits asymptotic optimality properties.

Empirically, we show that our method preserves fairness well for several base stochastic classifiers and datasets. We also show that it constitutes a new competitive avenue to learn fair deterministic decision models.

8.7 Federated and Decentralized Learning

Participants: Batiste Le Bars , Marc Damie , Aleksei Korneev , Marc Tommasi.

Fedivertex: a Graph Dataset based on Decentralized Social Networks for Trustworthy Machine Learning [26]

Decentralized machine learning - where each client keeps its own data locally and uses its own computational resources to collaboratively train a model by exchanging peer-to-peer messages - is increasingly popular, as it enables better scalability and control over the data. A major challenge in this setting is that learning dynamics depend on the topology of the communication graph, which motivates the use of real graph datasets for benchmarking decentralized algorithms. Unfortunately, existing graph datasets are largely limited to for-profit social networks crawled at a fixed point in time and often collected at the user scale, where links are heavily influenced by the platform and its recommendation algorithms. The Fediverse, which includes several free and open-source decentralized social media platforms such as Mastodon, Misskey, and Lemmy, offers an interesting real-world alternative. We introduce Fedivertex, a new dataset of 182 graphs, covering seven social networks from the Fediverse, crawled weekly over 14 weeks. We release the dataset along with a Python package to facilitate its use, and illustrate its utility on several tasks, including a new defederation task, which captures a process of link deletion observed on these networks.

A Survey on Verifiable Cross-Silo Federated Learning [20]

Federated Learning (FL) is a widespread approach that allows training machine learning (ML) models with data distributed across multiple storage units. In cross-silo FL, which often appears in domains like healthcare or finance, the number of participants is moderate, and each party typically represents a well-known

organization. For instance, in medicine data owners are often hospitals or data hubs which are well-established entities. However, malicious parties may still attempt to disturb the training procedure in order to obtain certain benefits, for example, a biased result or a reduction in computational load. While one can easily detect a malicious agent when data used for training is public, the problem becomes much more acute when it is necessary to maintain the privacy of the training dataset. To address this issue, there is recently growing interest in developing verifiable protocols, where one can check that parties do not deviate from the training procedure and perform computations correctly. In this paper, we present a survey on verifiable cross-silo FL. We analyze various protocols, fit them in a taxonomy, and compare their efficiency and threat models. We also analyze Zero-Knowledge Proof (ZKP) schemes and discuss how their overall cost in a FL context can be minimized. Lastly, we identify research gaps and discuss potential directions for future scientific work.

Adaptive collaboration for online personalized distributed learning with heterogeneous clients [49]

We study the problem of online personalized decentralized learning with N statistically heterogeneous clients collaborating to accelerate local training. An important challenge in this setting is to select relevant collaborators to reduce gradient variance while mitigating the introduced bias. To tackle this, we introduce a gradient-based collaboration criterion, allowing each client to dynamically select peers with similar gradients during the optimization process. Our criterion is motivated by a refined and more general theoretical analysis of the All-for-one algorithm, proved to be optimal in Even et al. (2022) for an oracle collaboration scheme. We derive excess loss upper-bounds for smooth objective functions, being either strongly convex, non-convex, or satisfying the Polyak-Łojasiewicz condition; our analysis reveals that the algorithm acts as a variance reduction method where the speed-up depends on a *sufficient variance*. We put forward two collaboration methods instantiating the proposed general schema; and we show that one variant preserves the optimality of All-for-one. We validate our results with experiments on synthetic and real datasets.

Federated Learning for MRI-based BrainAGE: a multicenter study on post-stroke functional outcome prediction [51]

Objective: Brain-predicted age difference (BrainAGE) is a neuroimaging biomarker reflecting brain health. However, training robust BrainAGE models requires large datasets, often restricted by privacy concerns. This study evaluates the performance of federated learning (FL) for BrainAGE estimation in ischemic stroke patients treated with mechanical thrombectomy, and investigates its association with clinical phenotypes and functional outcomes. **Methods:** We used FLAIR brain images from 1674 stroke patients across 16 hospital centers. We implemented standard machine learning and deep learning models for BrainAGE estimates under three data management strategies: centralized learning (pooled data), FL (local training at each site), and single-site learning. We reported prediction errors and examined associations between BrainAGE and vascular risk factors (e.g., diabetes mellitus, hypertension, smoking), as well as functional outcomes at three months post-stroke. Logistic regression evaluated BrainAGE's predictive value for these outcomes, adjusting for age, sex, vascular risk factors, stroke severity, time between MRI and arterial puncture, prior intravenous thrombolysis, and recanalisation outcome. **Results:** While centralized learning yielded the most accurate predictions, FL consistently outperformed single-site models. BrainAGE was significantly higher in patients with diabetes mellitus across all models. Comparisons between patients with good and poor functional outcomes, and multivariate predictions of these outcomes showed the significance of the association between BrainAGE and post-stroke recovery. **Conclusion:** FL enables accurate age predictions without data centralization. The strong association between BrainAGE, vascular risk factors, and post-stroke recovery highlights its potential for prognostic modeling in stroke care.

8.8 Conformal Prediction for Trustworthy Machine Learning

Participants: Batiste Le Bars.

On Volume Minimization in Conformal Regression [30]

We study the question of volume optimality in split conformal regression, a topic still poorly understood in comparison to coverage control. Using the fact that the calibration step can be seen as an empirical volume minimization problem, we first derive a finite-sample upper-bound on the excess volume loss of the interval returned by the classical split method. This important quantity measures the difference in length between the interval obtained with the split method and the shortest oracle prediction interval. Then, we introduce EffOrt, a methodology that modifies the learning step so that the base prediction function is selected in order to minimize the length of the returned intervals. In particular, our theoretical analysis of the excess volume loss of the prediction sets produced by EffOrt reveals the links between the learning and calibration steps, and notably the impact of the choice of the function class of the base predictor. We also introduce Ad-EffOrt, an extension of the previous method, which produces intervals whose size adapts to the value of the covariate. Finally, we evaluate the empirical performance and the robustness of our methodologies.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

We have started two new CIFRE contracts in 2023. We continue these collaborations until 2026.

Transfer learning for text anonymization

Participants: Damien Sileo, Marc Tommasi, Gabriel Loiseau.

VADE is a major company that processes emails at large scale to detect attacks like phishing.

In this project we design utility and privacy evaluation methods based on the combination of many tasks and objectives, relevant in the text (email) context. We study and compare approaches based on text generation or based on the replacement or obfuscation of selected entities, to tune the privacy utility trade-off.

Synthetic data generation with privacy constraints

Participants: Aurélien Bellet, Marc Tommasi, Clément Pierquin.

Craft.ai is a company whose activity was originally focused on explainable models for time series. It offers now MLops solutions based on AI with trustworthy guarantees. In this bilateral project with Craft.ai, Magnet brings expertise in privacy preserving machine learning for the generation of synthetic data.

The project is organized in four major axes. The definition of quality metrics for synthetic data; the design of algorithms for synthetic data generation with differential privacy guarantees; the definition of theoretical and empirical bounds on privacy associated with the release of synthetic data sets or generative models; some applications on time series or correlated data.

10 Partnerships and cooperations

10.1 International research visitors

10.1.1 Visits of international scientists

Inria International Chair Nicolas Papernot: AI Treaties

- Host teams: PreMeDICAL (primary host), Magnet, Privatics.

Participants: Nicolas Papernot, Marc Tommasi , Michaël Perrot , Raouf Kerkouche , Jan Ramon , Damien Sileo , Pascal Denis , Mikaela Keller , Batiste Le Bars.

The research agenda during the term of the Inria Chair will thus be structured around three major research thrusts, which will inform each other:

1. Towards differentially private decentralized learning with applications to healthcare.
2. How can learning algorithms be co-designed with cryptographic protocols to obtain verifiable certificates at the conclusion of training?
3. How will competition, cooperation, and coordination between countries, companies impact the stability of AI governance?

10.1.2 Visits to international teams

Research stays abroad

Participant: Dinh Viet-Toan Le.

Visited institution: McGill University

Country: Canada

Dates: August 15th 2025

Context of the visit: Presented his work to [Distributed Digital Music Archives and Libraries Lab](#)

Mobility program/type of mobility: paper accepted to IJCAI Conference in Montréal

10.2 European initiatives

10.2.1 Horizon Europe

TRUMPET

Participant: Jan Ramon (*contact person*).

Title: TRUStworthy Multi-site Privacy Enhancing Technologies

Duration: From October 1, 2022 to December 31, 2025

Partners:

- INRIA, France
- TIMELEX, Belgium
- Technovative Solutions LTD, United Kingdom
- Fundacion Centro Tecnoloxico de Telecomunicacions de Galicia (GRADIANT), Spain
- Commissariat à l'Énergie Atomique et aux Énergies alternatives (CEA), France
- Istituto Romagnolo per lo Studio dei Tumori Dino Amadori - IRST SRL (IRST), Italy
- Centre Hospitalier Universitaire de Liege (CHUL), Belgium

- Türkiye Cumhuriyeti Sağlık Bakanlığı (MOH), Türkiye
- Universidad de Vigo (UVIGO), Spain
- Arteevo Technologies LTD (ARTEEVO), Israel

Inria contact: Jan Ramon

Coordinator:

Summary: In recent years, Federated Learning (FL) has emerged as a revolutionary privacy-enhancing technology and, consequently, has quickly expanded to other applications.

However, further research has cast a shadow of doubt on the strength of privacy protection provided by FL. Potential vulnerabilities and threats pointed out by researchers included a curious aggregator threat; susceptibility to man-in-the-middle and insider attacks that disrupt the convergence of global and local models or cause convergence to fake minima; and, most importantly, inference attacks that aim to re-identify data subjects from FL's AI model parameter updates.

The goal of TRUMPET is to research and develop novel privacy enhancement methods for Federated Learning, and to deliver a highly scalable Federated AI service platform for researchers, that will enable AI-powered studies of siloed, multi-site, cross-domain, cross border European datasets with privacy guarantees that exceed the requirements of GDPR. The generic TRUMPET platform will be piloted, demonstrated and validated in the specific use case of European cancer hospitals, allowing researchers and policymakers to extract AI-driven insights from previously inaccessible cross-border, cross-organization cancer data, while ensuring the patients' privacy. The strong privacy protection accorded by the platform will be verified through the engagement of external experts for independent privacy leakage and re-identification testing.

A secondary goal is to research, develop and promote with EU data protection authorities a novel metric and tool for the certification of GDPR compliance of FL implementations.

The consortium is composed of 9 interdisciplinary partners: 3 Research Organizations, 1 University, 3 SMEs and 2 Clinical partners with extensive experience and expertise to guarantee the correct performance of the activities and the achievement of the results.

FLUTE

Participant: Jan Ramon (*contact person*).

Title: Federate Learning and mUlti-party computation Techniques for prostatE cancer

Duration: From May 1, 2023 to April 30, 2026

Partners:

- INRIA, France
- Quibim Sociedad Limitada (QUIBIM), Spain
- TIMELEX, Belgium
- Technovative Solutions LTD, United Kingdom
- HL7 Europe Foundation, Belgium
- Fundacion Centro Tecnoloxico de Telecomunicacions de Galicia (GRADIANT), Spain
- Siemens SRL, Romania
- Universitat Politècnica de Catalunya (UPC), Spain
- Istituto Romagnolo per lo Studio dei Tumori Dino Amadori - IRST SRL (IRST), Italy
- Centre Hospitalier Universitaire de Liege (CHUL), Belgium

- Fundacio Hospital Universitari Vall d'Hebron - Institut de Recerca (VHIR), Spain
- Arteevo Technologies LTD (ARTEEVO), Israel

Inria contact: Jan Ramon

Coordinator:

Summary: The FLUTE project will advance and scale up data-driven healthcare by developing novel methods for privacy-preserving cross-border utilization of data hubs. Advanced research will be performed to push the performance envelope of secure multi-party computation in Federated Learning, including the associated AI models and secure execution environments. The technical innovations will be integrated in a privacy-enforcing platform that will provide innovators with a provenly secure environment for federated healthcare AI solution development, testing and deployment, including the integration of real world health data from the data hubs and the generation and utilization of synthetic data. To maximize the impact, adoption and replicability of the results, the project will contribute to the global HL7 FHIR standard development, and create novel guidelines for GDPR-compliant cross-border Federated Learning in healthcare.

To demonstrate the practical use and impact of the results, the project will integrate the FLUTE platform with health data hubs located in three different countries, use their data to develop a novel federated AI toolset for diagnosis of clinically significant prostate cancer and perform a multi-national clinical validation of its efficacy, which will help to improve predictions of aggressive prostate cancer while avoiding unnecessary biopsies, thus improving the welfare of patients and significantly reducing the associated costs.

Team. The 11-strong consortium will include three clinical / data partners from three different countries, three technology SMEs, three technology research partners, a legal/ethics partner and a standards organization.

Collaboration. In accordance with the priorities set by the European Commission, the project will target collaboration, cross-fertilization and synergies with related national and international European projects.

10.3 National initiatives

10.3.1 ANR PMR (2020-2025)

Participants: Jan Ramon (*contact person*), Marc Tommasi.

Given the growing awareness of privacy risks of data processing, there is an increasing interest in privacy-preserving learning. However, shortcomings in the state of the art limit the applicability of the privacy-preserving learning paradigm. First, most approaches assume too optimistically a honest-but-curious setting. Second, most approaches consider one learning task in isolation, not accounting for the context where querying is a recurring activity. We will investigate new algorithms and models that address these shortcomings. Among others, (i) our algorithms will combine privacy-preserving properties of differential privacy with security offered by cryptography and (ii) based on models of information flows in integrated data handling processes, we will build more refined models analyzing the implications of repeated querying. We will demonstrate the utility of our new theory and algorithms by proposing strategies to realistically apply them in significant real-world problems illustrated through use cases in the medical domain

10.3.2 FedMalin. INRIA Defi (2021-2026)

Participants: Jan Ramon, Marc Tommasi (*contact person*), Michaël Perrot, Batiste Le Bars, Edwige Cyffers, Brahim Erraji, Luis Lugo, Paul Andrey.

In many use-cases of Machine Learning (ML), data is naturally decentralized: medical data is collected and stored by different hospitals, crowdsensed data is generated by personal devices, etc. Federated Learning (FL) has recently emerged as a novel paradigm where a set of entities with local datasets collaboratively train ML models while keeping their data decentralized.

FedMalin is a research project that spans 10 Inria research teams and aims to push FL research and concrete use-cases through a multidisciplinary consortium involving expertise in ML, distributed systems, privacy and security, networks, and medicine. We propose to address a number of challenges that arise when FL is deployed over the Internet, including privacy and fairness, energy consumption, personalization, and location/time dependencies.

FedMalin will also contribute to the development of open-source tools for FL experimentation and real-world deployments, and use them for concrete applications in medicine and crowdsensing.

10.3.3 COMANCHE: Computational Models of Lexical Meaning and Change. INRIA Action Exploratoire (2022-2026)

Participants: Pascal Denis (*contact person*), Mikaela Keller, Bastien Liétard, Nassim Boudjenah.

Comanche proposes to transfer and adapt recent Natural Language representation learning algorithms from deep learning to model the evolution of the meaning of words, and to confront these computational models to theories on language acquisition and the diachrony of languages. At the crossroads between machine learning, psycholinguistics and historical linguistics, this project will make it possible to validate or revise some of these theories, but also to bring out computational models that are more sober in terms of data and computations because they exploit new inductive biases inspired by these disciplines.

In collaboration with UMR SCALAB (CNRS, Université de Lille), l'Unité de Recherche STIH (Sorbonne Université), et l'UMR ATILF (CNRS, Université de Lorraine).

10.3.4 IPoP, Projet interdisciplinaire sur la protection des données personnelles, PEPR Cybersécurité (2022-2028).

Participants: Jan Ramon, Marc Tommasi (*contact person*), Michaël Perrot, Raouf Kerkouche, Batiste Le Bars, Paul Andrey, Jean Dufraiche, Shreya Venugopal.

Digital technologies provide services which can greatly increase quality of life (e.g. connected e-health devices, location based services, or personal assistants). However, these services can also raise major privacy risks, as they involve personal data, or even sensitive data. Indeed, this notion of personal data is the cornerstone of French and European regulations, since processing such data triggers a series of obligations that the data controller must abide by. This raises many multidisciplinary issues, as the challenges are not only technological, but also societal, judiciary, economic, political and ethical.

The objectives of this project are thus to study the threats on privacy that have been introduced by these new services, and to conceive theoretical and technical privacy-preserving solutions that are compatible with French and European regulations, that preserve the quality of experience of the users. These solutions will be deployed and assessed, both on the technological and legal sides, and on their societal acceptability. In order to achieve these objectives, we adopt an interdisciplinary approach, bringing together many diverse fields: computer science, technology, engineering, social sciences, economy and law.

The project's scientific program focuses on new forms of personal information collection, on Artificial Intelligence (AI) and its governance, data anonymization techniques, personal data management and distributed calculation protocol privacy preserving infrastructures, differential privacy, personal data legal protection and compliance, and all the associated societal and ethical considerations. This unifying interdisciplinary research program brings together internationally recognized research teams (from universities, engineering schools and institutions) working on privacy, and the French Data Protection Authority (CNIL).

This holistic vision of the issues linked to personal data protection will on the one hand let us propose solutions to the scientific and technological challenges and on the other help us confront these solutions in many different ways, in the context of interdisciplinary collaborations, thus leading to recommendations and proposals in the field of regulations or legal frameworks. This comprehensive consideration of all the issues aims at encouraging the adoption and acceptability of the solutions proposed by all stakeholders, legislators, data controllers, data processors, solution designers, developers all the way to end-users.

10.3.5 SSF-ML-DH, PEPR Santé Numérique (2022-2028).

Participants: Marc Tommasi (*contact person*), Batiste Le Bars.

The healthcare sector (public and private) generates an unparalleled amount of data from sources as diverse as electronic medical records, advanced imaging techniques, high-throughput sequencing, wearable devices, and public health data. Leveraging these massive datasets through sophisticated machine learning algorithms has the potential to transform medical practice by enabling the development of more effective and personalized treatments, interventions, and public policies, ultimately improving healthcare delivery and population well-being. However, the highly sensitive nature of health data, cybersecurity risks, biases in the data, and the lack of robustness in machine learning algorithms are key obstacles currently preventing the full realization of recent advancements in artificial intelligence.

To overcome these challenges, it is essential to address ethical, legal, security, and robustness issues. This project aims to develop new machine learning algorithms that account for the multi-scale and heterogeneous characteristics of health data while ensuring privacy, robustness against adversarial attacks and changes in data and model dynamics, and fairness for underrepresented populations. By addressing these obstacles, we hope to unlock the barriers that hinder the deployment of innovative solutions in digital health.

Specifically, the project will focus on the following challenges: (i) privacy-preserving learning through differential privacy techniques and homomorphic encryption; (ii) federated learning by balancing accuracy and privacy; (iii) robustness against adversarial attacks and changes in data and model dynamics; (iv) automated “forgetting” mechanisms to implement the right to be forgotten.

The project, part of Project of PEPR Digital Health (2023-2027), brings together a unique consortium of experts in machine learning, cybersecurity, statistics, and medical applications. Moreover, it is strategically positioned between two national programs (PEPRs: Cybersecurity and Digital Health), providing a unique opportunity to disseminate knowledge and best practices.

10.3.6 CAPS’UL (2023-2028)

Participant: Marc Tommasi (*contact person*), Paul Andrey, Simon Decomble.

The project is built around 3 axes.

1. Promote a common digital health culture among all current and future healthcare professionals: cybersecurity issues, legal and ethical regulation of healthcare data, communication and digital health tools, telehealth framework.
2. Design a high-performance tool for practical situations, enabling concrete and effective collaboration between the various training, socio-economic and medico-social players in the implementation of training courses. This shared resource center will provide a credible immersive environment (real software and simulated healthcare data) and teaching scenarios for the entire teaching community. Co-constructed with industry software publishers, it will be accessible from simulation centers and remotely, to meet the different needs of the region.
3. Train professionals in the new digital health support professions, by emphasizing the delivery of “health and specific digital issues” courses that are shared between the various existing courses. These innovative, coherent schemes will serve as demonstrators of excellence on a regional scale.

Magnet will provide tools for synthetic data generation with privacy guarantees dedicated to the immersive environment.

10.3.7 ANR-JCJC FaCTor: Fairness Constraints and Guarantees for Trustworthy Machine Learning (2023-2027)

Participants: Michaël Perrot (*contact person*), Marc Tommasi, Shreya Venugopal.

The goal of the FaCTor project is to provide ML practitioners with theoretically well founded means to develop algorithms that come with fairness guarantees. It points toward the development of trustworthy and socially acceptable ML solutions. The end goal is to make the models more accountable and in line with the requirements of the law, ensuring that the benefits of ML are not limited to a subset of the population.

10.3.8 REDEEM: Resilient, Decentralized and Privacy-Preserving Machine Learning, PEPR IA (2022-2028).

Participants: Jan Ramon (*contact person*), Marc Tommasi, Michaël Perrot, Arnaud Descours, Batiste Le Bars, Raouf Kerkouche.

The vision of distributed AI is attractive because it contributes to user empowerment by limiting the dissemination of personal and confidential information to a single node in the network and it makes systems independent of a superior force that would decide what is good for everyone. But on the other hand it opens up major issues of security and robustness: how can we guarantee the compliance of a model learned in another context? How can we protect our AI network from the introduction of biased knowledge, malicious or not, or even “backdoor” functions? If the pooling consists of a simultaneous optimisation, how can we ensure the validity of contributions that are not always explicable?

The action led on the theme of distributed AI is therefore at the confluence of the topics Embedded and Frugality (distributed systems are frequently low-resource embedded systems such as telephones, vehicles or autonomous robots) and Trust, as the issues of security, reliability and robustness are shed in a new light in collaborative AI.

The REDEEM project brings together a consortium of complementary teams and researchers, with primary expertise in machine learning, distributed optimization, consensus algorithms and game theory. It also associates a unique spectrum of research orientation, from highly theoretical work on convergence of distributed learning algorithms to extensive experiences towards practical and efficient implementations as well as innovative dissemination activities.

10.3.9 ANR-JCJC Adada: Adada: Adaptive Datasets for Enhancing Reasoning in Large Language Models (2024-2028)

Participants: Damien Sileo (*contact person*), Pascal Denis.

Large Language Models (LLMs) are neural networks designed for text completion, playing a pivotal role in various Natural Language Processing (NLP) applications such as conversational assistance and document analysis. Given the widening scope of LLM applications, generating truly useful completions goes beyond mere linguistic fluency. It requires logical precision, multi-step reasoning abilities, and adherence to user-defined constraints. These capabilities are essential for tackling the implicit reasoning tasks woven into everyday scenarios, from interpreting texts with embedded rules to evaluating products against technical specifications or identifying inconsistencies. At their core, such tasks involve complex logical problems intertwined with the nuances of natural language and with background knowledge. Logical reasoning remains challenging for current LLMs. As a countermeasure, we can train neural models to mimic the output of

symbolic reasoning systems (e.g., logic theorem provers, or other algorithms) on procedurally generated problems, like Q which actually comes from the RuleTaker dataset, to sharpen their reasoning capabilities. This training improves accuracy on human-authored problems. However synthetic problem datasets are currently generated once, sometimes not reproducibly. They can quickly become too easy for ongoing models after being included in the training data or due to model scaling. Adada proposes a novel framework to distill modern symbolic reasoning into language models through evolutive synthetic datasets. By explicitly steering problem generation to improve a specific model on a targeted downstream task, Adada seeks to continuously enhance language models for reasoning-intensive applications such as technical documentation understanding, commonsense reasoning, and legal analysis.

10.3.10 ANR Melissa: METHodological contributions in statistical Learning InSpired by SurfAce engineering (2025-2029)

Participants: Marc Tommasi [contact person] , Batiste Le Bars , Rémi Gilleron , Jan Ramon.

The underlying dynamics of many physical problems are governed by parameterized partial differential equations (PDEs). Despite important scientific advances in numerical simulation, solving efficiently PDEs remains complex and often prohibitively expensive. Physics-informed Machine Learning (PiML) has recently emerged as a promising way to learn efficient surrogate solvers, and augment the physical laws by leveraging knowledge extracted from data. From a machine learning perspective, ignoring the fundamental principles of the underlying physics may lead to ill-posed problems and thus to implausible solutions yielding poor generalization.

Numerous algorithmic contributions in deep-learning have recently exploited domain knowledge for (i) designing suitable physics-regularized loss functions, (ii) initializing neural networks with meaningful parameters, (iii) guiding the design of consistent architectures, or (iv) building hybrid models.

Despite indisputable advances, PiML remains an emerging topic with several open problems that remain to be addressed: (i) Deriving generalization/approximation guarantees; (ii) Learning with a limited amount of data; (iii) Augmenting partially known physical laws; (v) Modeling uncertainty; (vi) Building foundation models for physics.

Developing suited solutions that tackle these interrelated challenges is crucial for the usability of PiML in realistic scenarios. This is the goal of MELISSA which gathers 3 teams (Inria MALICE, Inria MAGNET and MLIA) with a strong expertise at the interface of machine learning, optimization and physics. By conducting this project from both theoretical and algorithmic perspectives, the objective is to design the next generation of provably accurate PiML algorithms in the challenging context of laser-matter interaction where data is scarce and the available physical laws only partially explain the observed dynamics.

10.4 Regional initiatives

10.4.1 Cross Disciplinary Project (CDP) Prime Next Gen: NEXT-GENERATION PRECISION medicine in Inflammatory and METabolic diseases (2026-2030)

Participants: Marc Tommasi [contact person] , Jan Ramon , Michaël Perrot.

This project has been accepted in 2025 and will start in 2026.

The overall objective of PRIME NEXT-GEN is to define, characterize and validate precise endotypes of patients with metabolic diseases and IMIDs, towards precision medicine. The scientific hypotheses are that 1. the level of meta-inflammation will be key in defining these endotypes 2. studying metabolic diseases and IMIDs together could provide a new clinically relevant pathophysiological nosography of these associated diseases. The further innovative approach developed in this project is to validate these endotypes at all pathophysiological, temporal and acceptability levels.

10.4.2 Cross Disciplinary Project (CDP) LOOP: closed Loop neurotechnologies: from sensors to applications (2026-2030)

Participants: Marc Tommasi [contact person] , Jan Ramon.

This project has been accepted in 2025 and will start in 2026.

the project will implement an interdisciplinary strategy applied to the development of a mechanism-based therapy for drug-resistant hallucinations in patients with schizophrenia. The project will be structured around three main research pillars: 1) the algorithmic level of closed-loop systems (online signal processing, machine learning); 2) the design of a complete and non-invasive solution to address the question of refractory hallucinations, and study how closed-loop interaction schemes can be beneficial for highly impaired psychiatric patients; 3) the design of the sensors themselves and translational research with animal models, in order to enhance the quality of the recordings and pave the way for future invasive miniaturized solutions. Finally, in addition to these three vertical research axes, this CDP also seeks to offer a unique opportunity to examine the development of neurotechnologies in the accelerating context of regulations at the European level, with law and ethics forming the first cross-cutting axis.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: selection

- Marc Tommasi served as Area Chair for UAI and ICLR, PC member of ECML, APVP and CAP.
- Jan Ramon served as reviewer for AAAI, AISTATS, ECML-PKDD, ICLR, ICML, IJCAI, NeurIPS, SDM, UAI and Icbimb@ICLR.
- Damien Sileo served as a reviewer for ACL.
- Michaël Perrot served as reviewer for ICML, NeurIPS, AISTATS, and CAP.
- Batiste Le Bars served as a reviewer for NeurIPS, AISTATS and ICLR
- Pascal Denis was Action Editor for ACL Rolling Review, and served as PC member for COLING 2025
- Mikaela Keller served as a reviewer for ACL, AISTATS and CAP

11.1.2 Journal

Member of editorial boards

- Jan Ramon is member of the editorial boards of Machine Learning Journal (MLJ), Data Mining and Knowledge Discovery (DMKD).
- Pascal Denis standing reviewer for Transactions of the Association for Computational Linguistics (TACL).

Reviewer - reviewing activities

- Batiste Le Bars served as reviewer for the Electronic Journal of Statistics (EJS) and the SIAM Journal on Optimization (SIOPT)

11.1.3 Invited talks

- Jan Ramon gave presentations at IPEAC Secured/FLUTE workshop (Barcelona, 1/25); HS Booster final event (Brussels, 5/3/25); Data health summit (Brussels, 20/3/25); IHI-HaDEA meeting on synthetic data (online) : Synthetic data in the FLUTE project (03/25); pepr-ipop workshop on legal aspects of AI (Paris, 20/3/25); Privacy symposium (Venice, 20/3/25); FHIN-FLUTE meeting (online): introduction to FLUTE for the FHIN (www.fhin.be) consortium (05/25); EBDVF (Copenhagen); OncoLille (Lille, 24/11/25); CRIStAL axe IA (Lille, 19/12/25).
- Marc Tommasi gave presentations at FedMalin (Rennes, 02/25); the prospective INRIA seminar (Rungis, 03/25); RedChainLab workshop (Lyon, 06/25).
- Michaël Perrot gave a presentation in the SIGMA Team (Lille, 05/25).
- Damien Sileo gave a presentation at the LLM4Roq lecture group (Paris, 17/10/25)
- Batiste Le Bars gave a presentation in the PreMeDiCaL Team (Montpellier, 04/25); at the ANR Melissa Kick-off meeting (Saint-Etienne, 05/25); and during the Journées de Statistique de la SFdS (Marseille, 06/25)
- Raouf Kerkouche gave a presentation at PrivateAIM on Differentially private federated learning for localized control of infectious disease dynamics (Germany, 10/25)

11.1.4 Leadership within the scientific community

- Jan Ramon was member of the bureau of the Société Savante Francophone d'Apprentissage Machine (SSFAM)
- Pascal Denis is co-head of the CNRS GDR "Langues et langage à la croisée des disciplines" (LLcD) and co-animator of the Working Group NLP & Cognition of the CNRS GDR-TAL.

11.1.5 Scientific expertise

- Marc Tommasi was a member (scientific expert) of the recruitment committee of full professors at Saint-Etienne and Lille.
- Marc Tommasi was a member (scientific expert) of the recruitment committee of associate professors at Lille.
- Marc Tommasi was expert for CY Generation
- Marc Tommasi was member of the ANR CE 39, Sécurité Globale.
- Jan Ramon was reviewer for COST project proposals.
- Pascal Denis was a member (scientific expert) of the CRCN/ISFP recruitment committee at INRIA Center of Bordeaux University, as well as acting parity and equal opportunities co-officer on that committee.
- Mikaela Keller was a member (scientific expert) of recruitment committees of assistant professors in Besançon and Bordeaux

11.1.6 Research administration

- Mikaela Keller was a vice-president of CER (Commission Emploi Recherche) in the INRIA Center of Lille University and a facilitator of the CRIStAL-wide AI Axis that promotes discussion among the CRIStAL teams working on AI.
- Marc Tommasi is co-head of the DatInG group (3 teams, about 80 persons), member of the Conseil Scientifique du laboratoire CRIStAL and member of the Commission mixte CRIStAL/Faculty of Science, Lille University. He is member of the BCEP (bureau du comité des équipes projet).

- Pascal Denis is also a member of the network "référents données" at Inria and Université de Lille (Lille Open Research Data). He is also administrator of Inria membership to Linguistic Data Consortium (LDC). He is also member of the local sustainable development committee (CLDD) at INRIA Center of Lille University.
- Michaël Perrot is a substitute member of the Comité de Centre (named, representing the administration). Michaël Perrot is the local correspondent for Activity Reports in the Inria center at the university of Lille. Michaël Perrot is volunteer in the local AGOS team.
- Jan Ramon is a member of the Comité de Centre and member of the bureau of the SSFAM (Société Savante Francophone d'Apprentissage Machine)
- Damien Sileo is a member of the Comité de l'Evaluation de l'IA.
- Batiste Le Bars is in charge of the Magnet Seminar organization

11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

11.2.1 Teaching

- Licence MIASHS: Mikaela Keller, Data Science, 24h, L2, Université de Lille.
- Licence Informatique: Marc Tommasi, Introduction to AI, 24h, L2, Université de Lille.
- Master Computer Science: Mikaela Keller, Apprentissage profond, 24h, M1, Université de Lille.
- Master Computer Science: Mikaela Keller, Machine learning pour le traitement automatique du langage naturel, 24h, M2, Université de Lille.
- Master Computer Science: Marc Tommasi, Data Science, 48h, M1, Université de Lille.
- Master Computer Science: Batiste Le Bars, Data Science, 36h, M1, Université de Lille.
- Master Data Science: Marc Tommasi Seminars 24h.
- Master Data Science: Damien Sileo, Natural Language Processing, 24h, M2, Université de Lille et Ecole Centrale de Lille.
- Master Data Science: Damien Sileo, Natural Language Processing, 4.5h, M2, Institut Mines-Télécom.
- Master Data Science: Michaël Perrot, Fairness in Trustworthy Machine Learning, 24h, M2, Université de Lille et Ecole Centrale de Lille.
- Master Cognitive Sciences: Michaël Perrot, Fairness in Machine Learning, 6h, M2, Université de Lille.

11.2.2 Supervision

- Postdoc: Arnaud Descours. On federated optimization with lower communication cost. Since Nov. 2023. Jan Ramon
- Postdoc: Luis Lugo. On federated learning with energy budgets, since Jun. 24. Marc Tommasi and Romain Rouvoy.
- Postdoc: Jean-Baptiste Fermanian. On Personalized Federated Learning, since Oct. 25. Aurélien Bellet and Batiste Le Bars.
- Phd defended: Marc Damie. Secure protocols for verifiable decentralized machine learning. Jan Ramon with Andreas Peter (U. Twente, NL & U. Oldenburg, DE) Florian Hahn (University of Twente, NL).[40]
- Phd defended: Dinh Viet-Toan Le. Natural Language Processing approaches in the musical domain : suitability, performance and limits. Mikaela Keller and LOUIS BIGO [41]

- PhD defended: Antoine Barczewski. Transparent privacy-preserving machine learning. Jan Ramon.[39]
- PhD in progress: Bastien Liétard. Computational Models of Lexical Semantic Change, since Nov. 2022. ANNE CARLIER (Université Paris Sorbonne), Pascal Denis and Mikaela Keller
- PhD in progress: Aleksei Korneev. Trustworthy multi-site privacy-enhancing technologies, since Dec. 2022. Jan Ramon
- PhD in progress: Clément Pierquin. Synthetic data generation with privacy constraints, since Sept. 2023, Aurélien Bellet and Marc Tommasi
- PhD in progress: Gabriel Loiseau. Transfert and multitask learning approaches for text anonymization, since Sept. 2023, Damien Sileo and Marc Tommasi
- PhD in progress: Brahim Erraji. Fairness in Federated Learning, since Sept. 2023, Aurélien Bellet, CATUSCIA PALAMIDESSI and Michaël Perrot.
- PhD in progress: Shreya Venugopal. Guaranteed Fairness in Machine Learning, since Oct. 24, Michaël Perrot.
- PhD in progress: Jean Dufrêche. Fairness and Privacy in Machine Learning, since Oct. 24, Michaël Perrot, Marc Tommasi.
- PhD in progress: Thomas Boudou. Private and Byzantine-Robust Federated Learning, since Oct. 24, Aurélien Bellet and Batiste Le Bars
- PhD in progress: Paul Andrey, Synthetic data and privacy, since Nov 24. Marc Tommasi and Batiste Le Bars
- PhD in progress: Quentin Sinh. Towards a generic, decentralized, secure and privacy preserving automated learning, Jul 24, Jan Ramon
- PhD in progress: Dimitri Kachler. Methodologies for the Generation, Analysis and Selection of Procedurally Generated Data for the Training and Evaluation of Large Language Models, since Nov. 25, Damien Sileo and Pascal Denis
- PhD in progress: Nassim Boudjenah. Computational Models of Semantic Memory, since Dec. 25, Pascal Denis and Rémi Gilleron
- Engineer: Jules Boulet, FLUTE and TRUMPET projects, until May 25, Jan Ramon
- Engineer: Elina Thibeau Sutre, FLUTE and TRUMPET projects until June 25, Jan Ramon
- Engineer: Jules Yvon, FLUTE and TRUMPET projects, since Sep. 24, Jan Ramon
- Engineer: Younes Ikli, FLUTE and TRUMPET projects, since Sep. 24, Jan Ramon
- Engineer: Léonard Deroose. Development of TRUMPET privacy-preserving platform components, Since Sep. 2023. Jan Ramon.
- Engineer: Baptiste Cottier. Zero-knowledge verification of computations, since Jan. 2025. Jan Ramon.
- Engineer: Valentin Lacombe. Enhancing the core reasoning capabilities of LLMs with RLVR (Reinforcement Learning with Verifiable Reward), leveraging a modular problem generation library, since Feb. 2025. Damien Sileo.
- Engineer: Zakaria El Bouchouari, Privacy preserving decentralized learning prototypes for FLUTE and REDEEM, since Feb. 25, Jan Ramon
- Engineer: Alexandre Louvet, differential privacy and the security of multi-party computation in the FLUTE and TRUMPET projects, since Sept. 25, Jan Ramon
- Engineer: Simon Decomble. Development of synthetic data generators for the CAPS'UL project, since Apr. 2025. Jan Ramon.

11.2.3 Juries

- Marc Tommasi was member of the following PhD juries: Volodimir Mitarchuk (Rapporteur), Marianne Abi Kanaan (Rapporteur), Yacine Belal (Examineur), Louis Roussel (President), Viet-Toan Le (Directeur), Jade Garcia-Bourrée (Rapporteur), Pierre Jobic (Rapporteur).
- Marc Tommasi was president of the HDR of Charlotte Laclau.

11.2.4 Educational and pedagogical outreach

- Marc Tommasi is directeur des études for the Machine Learning master of Computer Science.

11.3 Popularization

11.3.1 Participation in Live events

- Michaël Perrot gave an introductory presentation to machine learning during a Journée du Numérique on “Intelligence artificielle : enjeux éducatifs et pistes pédagogiques” organized by INSPÉ Lille HdF (Arras, 03/25).

11.3.2 Others science outreach relevant activities

- Marc Tommasi gave a presentation on responsible AI at the "séminaire national des secrétaires généraux d'académies et de région académiques" (Roubaix, 6/25).
- Michaël Perrot lead the scientific team advising the participants of the Serious Game Jam on Artificial Intelligence organized by INSPÉ Lille HdF and the Université de Lille (Lille, 01/25).
- Michaël Perrot gave an introductory presentation to fairness in machine learning to SNT and NSI teachers for the Académie de Lille (Lille, 06/25).
- Damien Sileo was interviewed by [Episloon](#)
- Michaël Perrot participated to a round table on artificial intelligence during the Semaine NSI (Lille, 12/25).
- Batiste Le Bars gave a poster presentation at the AI Action Summit Conference (École Polytechnique, 02/25)
- Mikaela Keller co-organized with MATHIEU GIRAUD from Algomus, CRISAL, a one day seminar [journée de recherche en musiques](#) to encourage collaborations between several disciplines working with Music in Université de Lille (Lille, 03/25)
- Mikaela Keller participated to a round table with jurists on AI for research in law during Séminaire doctoral du CRDP (Lille, 07/25)

12 Scientific production

12.1 Major publications

- [1] A. Bellet, R. Guerraoui and H. Hendrikx. ‘Who started this rumor? Quantifying the natural differential privacy guarantees of gossip protocols’. In: *DISC 2020 - 34th International Symposium on Distributed Computing*. Freiburg / Virtual, Germany, Oct. 2020. URL: <https://hal.inria.fr/hal-02166432>.
- [2] A. Bellet, R. Guerraoui, M. Taziki and M. Tommasi. ‘Personalized and Private Peer-to-Peer Machine Learning’. In: *AISTATS 2018 - 21st International Conference on Artificial Intelligence and Statistics*. Lanzarote, Spain, Apr. 2018, pp. 1–20. URL: <https://hal.inria.fr/hal-01745796>.

- [3] M. Dehouck and P. Denis. ‘Delexicalized Word Embeddings for Cross-lingual Dependency Parsing’. In: *EACL*. Vol. 1. EACL 2017. Valencia, Spain, Apr. 2017, pp. 241–250. DOI: [10.18653/v1/E17-1023](https://doi.org/10.18653/v1/E17-1023). URL: <https://hal.inria.fr/hal-01590639>.
- [4] M. Dehouck and P. Denis. ‘Phylogenetic Multi-Lingual Dependency Parsing’. In: *NAACL 2019 - Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, United States, June 2019. URL: <https://hal.archives-ouvertes.fr/hal-02143747>.
- [5] P. Kairouz, B. H. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings et al. ‘Advances and Open Problems in Federated Learning’. In: *Foundations and Trends in Machine Learning* 14.1-2 (2021), pp. 1–210. URL: <https://hal.inria.fr/hal-02406503>.
- [6] O. Kuzelka, Y. Wang and J. Ramon. ‘Bounds for Learning from Evolutionary-Related Data in the Realizable Case’. In: *International Joint Conference on Artificial Intelligence (IJCAI)*. Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) 2016. New York, United States, July 2016. URL: <https://hal.archives-ouvertes.fr/hal-01422033>.
- [7] E. Lassalle and P. Denis. ‘Joint Anaphoricity Detection and Coreference Resolution with Constrained Latent Structures’. In: *AAAI Conference on Artificial Intelligence (AAAI 2015)*. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015). Austin, Texas, United States, Jan. 2015. URL: <https://hal.inria.fr/hal-01205189>.
- [8] B. Le Bars, A. Bellet, M. Tommasi, K. Scaman and G. Neglia. ‘Improved Stability and Generalization Guarantees of the Decentralized SGD Algorithm’. In: *ICML 2024 - The Forty-first International Conference on Machine Learning*. Vienne, Austria, 21st July 2024. URL: <https://hal.science/hal-04611418>.
- [9] S. Longpre, R. Mahari, A. Chen, N. Obeng-Marnu, D. Sileo, W. Brannon, N. Muennighoff, N. Khazam, J. Kabbara, K. Perisetla, X. Wu, E. Shippole, K. Bollacker, T. Wu, L. Villa, S. Pentland and S. Hooker. ‘A large-scale audit of dataset licensing and attribution in AI’. In: *Nature Machine Intelligence* 6.8 (30th Aug. 2024), pp. 975–987. DOI: [10.1038/s42256-024-00878-8](https://doi.org/10.1038/s42256-024-00878-8). URL: <https://hal.science/hal-04749695>.
- [10] G. Maheshwari, P. Denis, M. Keller and A. Bellet. ‘Fair NLP Models with Differentially Private Text Encoders’. In: *Findings of the Association for Computational Linguistics: EMNLP 2022*. Abu Dhabi, United Arab Emirates, 2022. URL: <https://hal.inria.fr/hal-03905094>.
- [11] P. Mangold, M. Perrot, A. Bellet and M. Tommasi. ‘Differential Privacy has Bounded Impact on Fairness in Classification’. In: *Proceedings of the 40th International Conference on Machine Learning*. International Conference on Machine Learning. Vol. 202. Honolulu, United States, 23rd July 2023. URL: <https://hal.science/hal-03902203>.
- [12] C. Pelekis, J. Ramon and Y. Wang. ‘H’older-type inequalities and their applications to concentration and correlation bounds’. In: *Indagationes Mathematicae* 28.1 (2017), pp. 170–182. DOI: [10.1016/j.indag.2016.11.017](https://doi.org/10.1016/j.indag.2016.11.017). URL: <https://hal.archives-ouvertes.fr/hal-01421953>.
- [13] C. Sabater, A. Bellet and J. Ramon. ‘An Accurate, Scalable and Verifiable Protocol for Federated Differentially Private Averaging’. In: *Machine Learning* (28th Oct. 2022). DOI: [10.1007/s10994-022-06267-9](https://doi.org/10.1007/s10994-022-06267-9). URL: <https://hal.inria.fr/hal-03820603>.
- [14] A. S. Shamsabadi, B. M. L. Srivastava, A. Bellet, N. Vauquier, E. Vincent, M. Maouche, M. Tommasi and N. Papernot. ‘Differentially private speaker anonymization’. In: *Proceedings on Privacy Enhancing Technologies* 2023.1 (1st Jan. 2023). URL: <https://hal.inria.fr/hal-03588932>.
- [15] B. M. L. Srivastava, N. Vauquier, M. Sahidullah, A. Bellet, M. Tommasi and E. Vincent. ‘Evaluating Voice Conversion-based Privacy Protection against Informed Attackers’. In: *ICASSP 2020 - 45th International Conference on Acoustics, Speech, and Signal Processing*. IEEE Signal Processing Society. Barcelona, Spain, May 2020, pp. 2802–2806. URL: <https://hal.inria.fr/hal-02355115>.

- [16] P. Vanhaesebrouck, A. Bellet and M. Tommasi. ‘Decentralized Collaborative Learning of Personalized Models over Networks’. In: *International Conference on Artificial Intelligence and Statistics (AISTATS)*. Fort Lauderdale, Florida., United States, Apr. 2017. URL: <https://hal.inria.fr/hal-01533182>.

12.2 Publications of the year

International journals

- [17] M. Bellegarda, G. Boddaert, S. Dufour, D. Knutsen and A. Brunelliere. ‘Neural evidence for perceiving a vowel merger after a social interaction within a native language’. In: *Brain and Language*. Brain and Language 261 (1st Feb. 2025), p. 105529. DOI: [10.1016/j.bandl.2024.105529](https://doi.org/10.1016/j.bandl.2024.105529). URL: <https://lilloa.hal.science/hal-05225371> (cit. on p. 15).
- [18] C. Bossaert, S. Volant, P. Andrey, V. Hedouin and E. Wiel. ‘Medicolegal obstacles in pre-hospital care: An overview of practice in the North of France in 2023’. In: *Archives of Legal Medicine*. Archives of Legal Medicine 16.4 (9th Dec. 2025), p. 200623. DOI: [10.1016/j.aolm.2025.200623](https://doi.org/10.1016/j.aolm.2025.200623). URL: <https://hal.science/hal-05512994>.
- [19] A. Fasquel, W. El Mardi, I. Bonnotte, D. Knutsen and A. Brunellière. ‘How does the creation of new semantic relationships during dialogue impact long-term semantic representations after dialogue?’ In: *Acta Psychologica* 260 (8th Sept. 2025), p. 105513. DOI: [10.1016/j.actpsy.2025.105513](https://doi.org/10.1016/j.actpsy.2025.105513). URL: <https://hal.science/hal-05442278> (cit. on p. 15).
- [20] A. Korneev and J. Ramon. ‘A Survey on Verifiable Cross-Silo Federated Learning’. In: *Transactions on Machine Learning Research Journal* (17th June 2025). URL: <https://hal.science/hal-05117141> (cit. on p. 24).
- [21] D.-V.-T. Le, L. Bigo, D. Herremans and M. Keller. ‘Natural Language Processing Methods for Symbolic Music Generation and Information Retrieval: a Survey’. In: *ACM Computing Surveys* 57.7 (6th Jan. 2025), pp. 1–40. DOI: [10.1145/3714457](https://doi.org/10.1145/3714457). URL: <https://hal.science/hal-04621444> (cit. on p. 18).
- [22] T. Leteno, M. Perrot, C. Laclau, A. Gourru and C. Gravier. ‘Fair Text Classification via Transferable Representations’. In: *Journal of Machine Learning Research* 26.239 (Dec. 2025), pp. 1–47. URL: <https://hal.science/hal-05099202> (cit. on p. 24).
- [23] L. Phan, A. Gatti, Z. Han, N. Li, J. Hu, H. Zhang, S. Shi, M. Choi, A. Agrawal, A. Chopra et al. ‘A benchmark of expert-level academic questions to assess AI capabilities’. In: *Nature* 649.8099 (28th Jan. 2026), pp. 1139–1146. DOI: [10.1038/s41586-025-09962-4](https://doi.org/10.1038/s41586-025-09962-4). URL: <https://hal.science/hal-04915593> (cit. on p. 14).
- [24] W. Sun, M. Li, D. Sileo, J. Davis and M.-F. Moens. ‘Generating Explanations in Medical Question-Answering by Expectation Maximization Inference over Evidence’. In: *ACM Transactions on Computing for Healthcare* 6.2 (2025), p. 23. DOI: [10.1145/3712296](https://doi.org/10.1145/3712296). URL: <https://hal.science/hal-05290617> (cit. on p. 15).

International peer-reviewed conferences

- [25] P. Andrey, B. Le Bars and M. Tommasi. ‘TAMIS: Tailored Membership Inference Attacks on Synthetic Data’. In: *Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2025, Porto, Portugal, September 15–19, 2025, Proceedings, Part V*. ECML PKDD 2025 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Vol. 16017. Lecture Notes in Computer Science. Porto (Portugal), Portugal: Springer Nature Switzerland, 27th Sept. 2025, pp. 203–220. DOI: [10.1007/978-3-032-06096-9_12](https://doi.org/10.1007/978-3-032-06096-9_12). URL: <https://hal.science/hal-05360487> (cit. on p. 22).
- [26] M. Damie and E. Cyffers. ‘Fedivertex: a Graph Dataset based on Decentralized Social Media’. In: *WWW '26 - ACM Web Conference 2026*. Dubai, United Arab Emirates, 2026. DOI: [10.1145/3774904.3792868](https://doi.org/10.1145/3774904.3792868). URL: <https://hal.science/hal-05317793> (cit. on p. 24).

- [27] M. Damie, J.-B. Leger, F. Hahn and A. Peter. ‘Revisiting the Attacker’s Knowledge in Inference Attacks Against Searchable Symmetric Encryption’. In: 23rd International Conference on Applied Cryptography and Network Security (ACNS 2025). Vol. 15826. Munich (Allemagne), Germany: Springer Nature Switzerland, 20th June 2025, pp. 370–399. doi: [10.1007/978-3-031-95764-2_15](https://doi.org/10.1007/978-3-031-95764-2_15). URL: <https://hal.science/hal-05317770> (cit. on p. 21).
- [28] B. van Dartel, M. Damie and F. Hahn. ‘Evaluating Membership Inference Attacks in Heterogeneous-Data Setups’. In: Applied Cryptography and Network Security Workshops. Vol. 15655. Lecture Notes in Computer Science. Munich, Germany: Springer Nature Switzerland, 25th Oct. 2026, pp. 109–117. doi: [10.1007/978-3-032-01823-6_7](https://doi.org/10.1007/978-3-032-01823-6_7). URL: <https://hal.science/hal-05340891> (cit. on p. 23).
- [29] D.-V.-T. Le and Y.-H. Yang. ‘METEOR: Melody-aware Texture-controllable Symbolic Orchestral Music Generation via Transformer VAE’. In: International Joint Conference on Artificial Intelligence AI, Arts & Creativity (IJCAI 2025). Montreal, Canada: International Joint Conferences on Artificial Intelligence Organization, 2025, pp. 10126–10134. doi: [10.24963/ijcai.2025/1125](https://doi.org/10.24963/ijcai.2025/1125). URL: <https://hal.science/hal-05228512> (cit. on p. 18).
- [30] B. Le Bars and P. Humbert. ‘On Volume Minimization in Conformal Regression’. In: *Proceedings of Machine Learning Research*. International Conference on Machine Learning (ICML). Vol. Volume 267: International Conference on Machine Learning, 13-19 July 2025, Vancouver Convention Center, Vancouver, Canada. Proceedings of Machine Learning Research. Vancouver, Canada, 13th July 2025. URL: <https://hal.science/hal-04945088> (cit. on p. 26).
- [31] G. Loiseau, D. Sileo, D. Riquet, M. Meyer and M. Tommasi. ‘TAROT: Task-Oriented Authorship Obfuscation Using Policy Optimization Methods’. In: Proceedings of the Sixth Workshop on Privacy in Natural Language Processing. Albuquerque, United States: Association for Computational Linguistics, Apr. 2025, pp. 14–31. doi: [10.18653/v1/2025.privatenlp-main.2](https://doi.org/10.18653/v1/2025.privatenlp-main.2). URL: <https://hal.science/hal-05299966> (cit. on p. 18).
- [32] G. Loiseau, D. Sileo, D. Riquet, M. Meyer and M. Tommasi. ‘Tau-Eval: A Unified Evaluation Framework for Useful and Private Text Anonymization’. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Suzhou, France: Association for Computational Linguistics, 2025, pp. 216–227. doi: [10.18653/v1/2025.emnlp-demos.16](https://doi.org/10.18653/v1/2025.emnlp-demos.16). URL: <https://hal.science/hal-05300069> (cit. on p. 17).
- [33] S. Longpre, N. Singh, M. Cherep, K. Tiwary, J. Materzynska, W. Brannon, R. Mahari, N. Obeng-Marnu, M. Dey, M. Hamdy, N. Saxena, A. Mustafa Anis, E. A. Alghamdi, V. Minh Chien, D. Yin, K. Qian, Y. Li, M. Liang, A. Dinh, S. Mohanty, D. Mataciunas, T. South, J. Zhang, A. N. Lee, C. S. Lund, C. Klamm, D. Sileo, D. Misra, E. Shippole, K. Klyman, L. James Validad Miranda, N. Muennighoff, S. Ye, S. Kim, V. Gupta, V. Sharma, X. Zhou, C. Xiong, L. Villa, S. Biderman, A. Pentland, S. Hooker and J. Kabbara. ‘BRIDGING THE DATA PROVENANCE GAP ACROSS TEXT, SPEECH, AND VIDEO’. In: ICLR 2025. Singapore (SG), Singapore, 24th Apr. 2025. URL: <https://hal.science/hal-05300062> (cit. on p. 14).
- [34] C. Pierquin, A. Bellet, M. Tommasi and M. Boussard. ‘Privacy Amplification Through Synthetic Data: Insights from Linear Regression’. In: ICML 2025 - 42nd International Conference on Machine Learning. Vancouver, Canada, 2025. doi: [10.48550/arXiv.2506.05101](https://doi.org/10.48550/arXiv.2506.05101). URL: <https://hal.science/hal-05237878> (cit. on p. 22).
- [35] N. Tomashenko, E. Vincent and M. Tommasi. ‘Analysis of Speech Temporal Dynamics in the Context of Speaker Verification and Voice Anonymization’. In: 2025 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2025). Hyderabad, India, 6th Apr. 2025. doi: [10.1109/ICASSP49660.2025.10887896](https://doi.org/10.1109/ICASSP49660.2025.10887896). URL: <https://hal.science/hal-04853872> (cit. on p. 19).
- [36] N. Tomashenko, E. Vincent and M. Tommasi. ‘Exploiting Context-dependent Duration Features for Voice Anonymization Attack Systems’. In: Interspeech 2025. Rotterdam, Netherlands, 18th Aug. 2025. URL: <https://hal.science/hal-05099074> (cit. on p. 19).

Conferences without proceedings

- [37] A. Gutiérrez Cisneros, A. Foucart and A. Brunellière. ‘Indirect Reply Processing in Multilingual Conversations when Inferring Speaker Meaning: an ERP study’. In: ISP 2025 - 17th International Symposium of Psycholinguistics. Barcelona, Spain, 26th May 2025. URL: <https://hal.science/hal-05436737> (cit. on p. 16).
- [38] D. Knutsen, W. S. Horton and A. Brunellière. ‘Faces and voices in dialogue: How partner-specific cues contribute to conversational memory’. In: The Annual Meeting of the Society for Text and Discourse. Padua, Italy, July 2025. URL: <https://hal.science/hal-05443306> (cit. on p. 17).

Doctoral dissertations and habilitation theses

- [39] A. Barczewski. ‘Enhancing Differentially private machine learning : Optimizations for Repeated Query scenarios’. Université de Lille, 15th Oct. 2025. URL: <https://theses.hal.science/tel-05458397> (cit. on pp. 22, 37).
- [40] M. Damie. ‘Privacy-Preserving Computations on Sparse Data’. Université de Lille, 11th Dec. 2025. URL: <https://hal.science/tel-05436190> (cit. on pp. 20, 36).
- [41] D.-V.-T. Le. ‘Modeling Symbolic Music with Natural Language Processing Approaches’. Université de Lille, 3rd Nov. 2025. URL: <https://hal.science/tel-05426752> (cit. on pp. 19, 36).

Reports & preprints

- [42] T. Boudou, B. L. Bars, N. Gupta and A. Bellet. *Generalization under Byzantine & Poisoning Attacks: Tight Stability Bounds in Robust Distributed Learning*. 2025. URL: <https://hal.science/hal-05245060> (cit. on p. 20).
- [43] M. Damie, F. Hahn, A. Peter and J. Ramon. *How to Securely Shuffle? A survey about Secure Shufflers for privacy-preserving computations*. 2025. DOI: [10.48550/arXiv.2507.01487](https://doi.org/10.48550/arXiv.2507.01487). URL: <https://hal.science/hal-05317803> (cit. on p. 21).
- [44] M. Damie, F. Hahn, A. Peter and J. Ramon. *Secure Sparse Matrix Multiplications and their Applications to Privacy-Preserving Machine Learning*. 2025. DOI: [10.48550/arXiv.2510.14894](https://doi.org/10.48550/arXiv.2510.14894). URL: <https://hal.science/hal-05319134> (cit. on p. 21).
- [45] M. Damie, F. Mazzone, F. Hahn, A. Peter and J. Ramon. *Noisy Function Secret Sharing and its applications to Differentially Private computations*. 31st Oct. 2025. URL: <https://hal.science/hal-05340908> (cit. on p. 21).
- [46] J.-B. Fermanian, B. Le Bars and A. Bellet. *Adaptive Personalized Federated Learning via Multi-task Averaging of Kernel Mean Embeddings*. 12th Feb. 2026. URL: <https://hal.science/hal-05506908>.
- [47] L. Györfi, P. Humbert and B. Le Bars. *Metric space valued Fréchet regression*. 4th Feb. 2026. URL: <https://hal.science/hal-05493499>.
- [48] V. Lacombe, V. Quesnel and D. Sileo. *Reasoning Core: A Scalable RL Environment for LLM Symbolic Reasoning*. 2025. DOI: [10.48550/arXiv.2509.18083](https://doi.org/10.48550/arXiv.2509.18083). URL: <https://hal.science/hal-05290611> (cit. on p. 13).
- [49] C. Philippenko, B. Le Bars, K. Scaman and L. Massoulie. *Adaptive collaboration for online personalized distributed learning with heterogeneous clients*. 9th July 2025. URL: <https://hal.science/hal-05175097> (cit. on p. 25).
- [50] V. Quesnel and D. Sileo. *Saturation-Driven Dataset Generation for LLM Mathematical Reasoning in the TPTP Ecosystem*. 2025. DOI: [10.48550/arXiv.2509.06809](https://doi.org/10.48550/arXiv.2509.06809). URL: <https://hal.science/hal-05290608> (cit. on p. 14).
- [51] V. Roca, M. Tommasi, P. Andrey, A. Bellet, M. Schirmer, H. Henon, L. Puy, J. Ramon, G. Kuchcinski, M. Bretzner and R. Lopes. *Federated Learning for MRI-based BrainAGE: a multicenter study on post-stroke functional outcome prediction*. 2025. DOI: [10.48550/arXiv.2506.15626](https://doi.org/10.48550/arXiv.2506.15626). URL: <https://inria.hal.science/hal-05244644> (cit. on p. 25).

- [52] C. Sabater, S. B. Mokhtar and J. Ramon. *Dropout-Robust Mechanisms for Differentially Private and Fully Decentralized Mean Estimation*. 4th June 2025. URL: <https://hal.science/hal-05308749> (cit. on p. 22).
- [53] D. Sileo. *Logic Haystacks: Probing LLMs Long-Context Logical Reasoning (Without Easily Identifiable Unrelated Padding)*. 2025. DOI: [10.48550/arXiv.2502.17169](https://doi.org/10.48550/arXiv.2502.17169). URL: <https://hal.science/hal-05300827> (cit. on p. 14).

Other scientific publications

- [54] A. Brunelliere, L. Ott, S. Kalenine and M. Pickering. ‘Interacting with someone shapes prediction in spoken-language comprehension’. In: *AMLaP 2025 - Architectures and Mechanisms for Language Processing*. Prague (République Tchèque), Czech Republic, 4th Sept. 2025. URL: <https://lilloa.hal.science/hal-05236859> (cit. on p. 16).
- [55] A. Gutiérrez Cisneros, A. Brunellière and A. Foucart. ‘Neural signature of indirect reply processing while listening to foreign accented dialogues’. In: *International Symposium on Bilingualism*. Donostia, Spain, 9th June 2025. URL: <https://hal.science/hal-05436740> (cit. on p. 17).
- [56] A. Gutiérrez Cisneros, A. Foucart and A. Brunellière. ‘Indirect Reply Processing in Multilingual Contexts: an ERP study’. In: *L3 WORKSHOP 2025: Multilingual Acquisition, Processing and Use*. Madrid, Spain, 3rd Oct. 2025. URL: <https://hal.science/hal-05436743> (cit. on p. 16).
- [57] D.-V.-T. Le, L. Bigo and M. Keller. *Evaluating Interval-based Tokenization for Pitch Representation in Symbolic Music Analysis*. 3rd Mar. 2025. URL: <https://hal.science/hal-04877659> (cit. on p. 19).