RESEARCH CENTRE

Inria Centre at the University of Lille

IN PARTNERSHIP WITH:

CNRS, Université de Lille

2024 ACTIVITY REPORT

Project-Team MAGNET

Machine Learning in Information Networks

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

DOMAIN

Perception, Cognition and Interaction

THEME

Data and Knowledge Representation and Processing

Inria

Contents

Project-Team MAGNET

Creation of the Project-Team: 2016 May 01

Keywords

Computer sciences and digital sciences

A5.7.3. – Speech

A5.8. – Natural language processing

A9.2. – Machine learning

A9.4. – Natural language processing

A9.9. – Distributed AI, Multi-agent

Other research topics and application domains

B2. – Health

B9.5.1. – Computer science

B9.5.6. – Data science

B9.6.8. – Linguistics

B9.6.10. – Digital humanities

B9.9. – Ethics

B9.10. – Privacy

1 Team members, visitors, external collaborators

Research Scientists

- Pascal Denis [INRIA, Researcher]
- Batiste Le Bars [INRIA, ISFP, from Aug 2024]
- Michael Perrot [INRIA, ISFP]
- Jan Ramon [INRIA, Senior Researcher]
- Damien Sileo [INRIA, ISFP]

Faculty Members

- Marc Tommasi [Team leader, UNIV LILLE, Professor]
- Angèle Brunelliere [UNIV LILLE, Professor Delegation, from Sep 2024]
- Rémi Gilleron [UNIV LILLE, Emeritus]
- Mikaela Keller [UNIV LILLE, Associate Professor Delegation]

Post-Doctoral Fellows

- Arnaud Descours [INRIA, Post-Doctoral Fellow]
- Vitalii Emelianov [INRIA, Post-Doctoral Fellow, until Jun 2024]
- Luis Eduardo Lugo Martinez [INRIA, Post-Doctoral Fellow, from Jun 2024]
- Imane Taibi [INRIA, Post-Doctoral Fellow, until Apr 2024]

PhD Students

- Paul Andrey [INRIA, from Nov 2024]
- Antoine Barczewski [UNIV LILLE]
- Moitree Basu [INRIA, until Feb 2024]
- Mouad Blej [INRIA, from Aug 2024]
- Ioan Tudor Cebere [INRIA, until Aug 2024]
- Edwige Cyffers [UNIV LILLE]
- Marc Damie [University of Twente]
- Jean Dufraiche [INRIA, from Oct 2024]
- Brahim Erraji [INRIA]
- Aleksei Korneev [UNIV LILLE]
- Dinh-Viet-Toan Le [UNIV LILLE]
- Bastien Lietard [INRIA]
- Gabriel Loiseau [VADE - HORNET, CIFRE]
- Gaurav Maheshwari [INRIA, until Mar 2024]

- Aymane Moataz [INRIA, from Jun 2024]
- Clement Pierquin [CRAFT.AI, CIFRE]
- Aurelien Said Housseini [INRIA]
- Quentin Sinh [INRIA, from Jul 2024]
- Shreya Venugopal [INRIA, from Oct 2024]

Technical Staff

- Paul Andrey [INRIA, Engineer, until Oct 2024]
- Mouad Blej [INRIA, Engineer, from Feb 2024 until Jul 2024]
- Jules Boulet [INRIA, Engineer, from Jun 2024]
- Leonard Deroose [INRIA, Engineer]
- Younes Ikli [INRIA, Engineer, from Sep 2024]
- Mou Li [INRIA, Engineer, until Mar 2024]
- Kevin Hubert N Gakosso [INRIA, Engineer, until Nov 2024]
- Quentin Sinh [INRIA, Engineer, until Jun 2024]
- Elina Thibeau-Sutre [INRIA, Engineer, from May 2024]
- Sophie Villerot [INRIA, Engineer, until Apr 2024]
- Jules Yvon [INRIA, Engineer, from Oct 2024]

Interns and Apprentices

- Martin Borquez Herrera [INRIA, Intern, from Mar 2024 until Jun 2024]
- Mathieu Dejoie [ENS RENNES, Intern, from May 2024 until Jul 2024]
- Laura Fuentes [UNIV LILLE, Intern, from Apr 2024 until Aug 2024]
- Achraf Laalou [INRIA, Intern, from Jun 2024 until Aug 2024]

Administrative Assistant

- Aurore Dalle [INRIA]

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 project FedMalin (see Section 10.2.2) is 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

We are proud of the results of the team's evaluation processes by both INRIA and HCERES.

7 New software, platforms, open data

7.1 New software

7.1.1 CoRTeX

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

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, practicioners 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

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

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.

News of the Year: A transfer contract was signed with the startup Nijta.

Contact: Nathalie Vauquier

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

7.2 Open data

FOL-NLI first order logic reasoning dataset

Contributors: Damien Sileo

Description: A dataset for logical reasoning, designed to test and improve language models' abilities to solve first-order logic (FOL) problems using a declarative grammar. The dataset includes procedurally generated problems, with semantic constraints and verbalizations in simplified English and TPTP theorem-proving language.

Dataset PID (DOI,...): 10.18653/v1/2024.emnlp-main.301

Project link: `https://huggingface.co/datasets/tasksource/FOL-nli`

Publications: Sileo, Damien. "Scaling Synthetic Logical Reasoning Datasets with Context-Sensitive Declarative Grammars." In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 2024.

Contact: Damien Sileo

Release contributions: Dataset, code and methodology for generating first-order logic problems with context-sensitive declarative grammars.

Data Provenance Initiative

Contributors: Damien Sileo, Shayne Longpre, Robert Mahari, Anthony Chen, and others.

Description: The Data Provenance Initiative (DPI) focuses on improving transparency in the datasets used for fine-tuning large language models. DPI addresses the neglect of metadata during dataset aggregation, recovering missing licenses and attributions, and tracing the origins of datasets. The initiative has uncovered inaccuracies in licenses on prominent platforms like GitHub and HuggingFace and has annotated datasets with additional details such as task types, geographical origins, and dates to enable comprehensive audits.

Dataset PID (DOI,...): 10.1038/s42256-024-00878-8

Project link: `https://www.dataprovenance.org/`

Publications: Longpre, Shayne, et al. "A large-scale audit of dataset licensing and attribution in AI." *Nature Machine Intelligence*, 2024. [18], [23]

Contact: Damien Sileo

Release contributions: DPI has developed tools for auditing dataset provenance, recovering licenses, and annotating datasets with metadata to support responsible AI development.

Tasksource

Contributors: Damien Sileo

Description: Tasksource is a project aimed at improving dataset usability for NLP tasks by offering standardized preprocessing frameworks. It provides a system for annotating preprocessing steps and has uploaded over a hundred datasets to HuggingFace. Tasksource harmonizes datasets for multi-task training and evaluation, offering one-line commands for loading datasets into standardized formats. It also provides reusable preprocessing annotations for easy alignment of datasets.

Dataset PID (DOI,...): 2024.lrec-main.1361

Project link: https://github.com/sileod/tasksource

Publications: Sileo, Damien. "tasksource: A Large Collection of NLP tasks with a Structured Dataset Preprocessing Framework." *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*.

Contact: Damien Sileo

Release contributions: Tasksource provides over 600 annotated task preprocessings and a backend to automate dataset alignment, facilitating multi-task training and evaluation.

8 New results

8.1 Natural Language Processing

Analyzing Byte-Pair Encoding on Monophonic and Polyphonic Symbolic Music: A Focus on Musical Phrase Segmentation [32]

3rd Workshop on NLP for Music and Audio (NLP4MusA) [32]

Byte-Pair Encoding (BPE) is an algorithm commonly used in Natural Language Processing to build a vocabulary of subwords, which has been recently applied to symbolic music. Given that symbolic music can differ significantly from text, particularly with polyphony, we investigate how BPE behaves with different types of musical content. This study provides a qualitative analysis of BPE's behavior across various instrumentations and evaluates its impact on a musical phrase segmentation task for both monophonic and polyphonic music. Our findings show that the BPE training process is highly dependent on the instrumentation and that BPE "supertokens" succeed in capturing abstract musical content. In a musical phrase segmentation task, BPE notably improves performance in a polyphonic setting, but enhances performance in monophonic tunes only within a specific range of BPE merges.

Consent in Crisis: The Rapid Decline of the AI Data Commons [23]

General-purpose artificial intelligence (AI) systems are built on massive swathes of public web data, assembled into corpora such as C4, RefinedWeb, and Dolma. To our knowledge, we conduct the first, large-scale, longitudinal audit of the consent protocols for the web domains underlying AI training corpora. Our audit of 14,000 web domains provides an expansive view of crawlable web data and how codified data use preferences are changing over time. We observe a proliferation of AI-specific clauses to limit use, acute differences in restrictions on AI developers, as well as general inconsistencies between websites' expressed intentions in their Terms of Service and their robots.txt. We diagnose these as symptoms of ineffective web protocols, not designed to cope with the widespread re-purposing of the internet for AI. Our longitudinal analyses show that in a single year (2023-2024) there has been a rapid crescendo of data restrictions from web sources, rendering 5%+ of all tokens in C4, or 28%+ of the most actively maintained, critical sources in C4, fully restricted from use. For Terms of Service crawling restrictions, a full 45% of C4 is now restricted. If respected or enforced, these restrictions are rapidly biasing the diversity, freshness, and scaling laws for general-purpose AI systems. We hope to illustrate the

emerging crises in data consent, for both developers and creators. The foreclosure of much of the open web will impact not only commercial AI, but also non-commercial AI and academic research.

Scaling Synthetic Logical Reasoning Datasets with Context-Sensitive Declarative Grammars [42]

Logical reasoning remains a challenge for natural language processing, but it can be improved by training language models to mimic theorem provers on procedurally generated problems. Previous work used domain-specific proof generation algorithms, which biases reasoning toward specific proof traces and limits auditability and extensibility. We present a simpler and more general declarative framework with flexible context-sensitive rules binding multiple languages (specifically, simplified English and the TPTP theorem-proving language). We construct first-order logic problems by selecting up to 32 premises and one hypothesis. We demonstrate that using semantic constraints during generation and careful English verbalization of predicates enhances logical reasoning without hurting natural English tasks. We use relatively small DeBERTa-v3 models to achieve state-of-the-art accuracy on the FOLIO human-authored logic dataset, surpassing GPT-4 in accuracy with or without an external solver by 12%.

MMAR: Multilingual and multimodal anaphora resolution in instructional videos [25]

Multilingual anaphora resolution identifies referring expressions and implicit arguments in texts and links to antecedents that cover several languages. In the most challenging setting, cross-lingual anaphora resolution, training data, and test data are in different languages. As knowledge needs to be transferred across languages, this task is challenging, both in the multilingual and cross-lingual setting. We hypothesize that one way to alleviate some of the difficulty of the task is to include multimodal information in the form of images (i.e. frames extracted from instructional videos). Such visual inputs are by nature language agnostic, therefore cross-and multilingual anaphora resolution should benefit from visual information. In this paper, we provide the first multilingual and multimodal dataset annotated with anaphoric relations and present experimental results for end-to-end multimodal and multilingual anaphora resolution. Given gold mentions, multimodal features improve anaphora resolution results by ~10% for unseen languages.

Natural Language Processing Methods for Symbolic Music Generation and Information Retrieval: a Survey [40]

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.

Towards an Onomasiological Study of Lexical Semantic Change through the Induction of Concepts [22]

Lexical Semantic Change, the temporal evolution of the mapping between word forms and concepts, can be studied under two complementary perspectives: semasiology studies how given words change in meaning over time, while onomasiology focuses on how some concepts change in how they are lexically realized. For the most part, existing NLP studies have taken the semasiological (i.e. word-to-concept) view. In this paper, we describe a novel computational methodology that takes an onomasiological (i.e.,

concept-to-word) view of semantic change by directly inducing concepts from word occurrences at the different time stamps. We apply our methodology to a French diachronic corpus. We examine the quality of obtained concepts and showcase how the results of our methodology can be used for the study of Lexical Semantic Change. We discuss its advantages and its early limitations.

To Word Senses and Beyond: Inducing Concepts with Contextualized Language Models [21]

Polysemy and synonymy are two crucial interrelated facets of lexical ambiguity. While both phenomena are widely documented in lexical resources and have been studied extensively in NLP,leading to dedicated systems, they are often being considered independently in practical problems. While many tasks dealing with polysemy (e.g. Word Sense Disambiguiation or Induction) highlight the role of word's senses,the study of synonymy is rooted in the study of concepts, i.e. meanings shared across the lexicon. In this paper, we introduce Concept Induction, the unsupervised task of learning a soft clustering among words that defines a set of concepts directly from data. This task generalizes Word Sense Induction. We propose a bi-level approach to Concept Induction that leverages both a local lemma-centric view and a global cross-lexicon view to induce concepts. We evaluate the obtained clustering on SemCor's annotated data and obtain good performance (BCubed F1 above 0.60). We find that the local and the global levels are mutually beneficial to induce concepts and also senses in our setting. Finally,we create static embeddings representing our induced concepts and use them on the Word-in-Context task, obtaining competitive performance with the State-of-the-Art.

Recipient Profiling: Predicting Characteristics from Messages [36]

It has been shown in the field of Author Profiling that texts may inadvertently reveal sensitive information about their authors, such as gender or age. This raises important privacy concerns that have been extensively addressed in the literature, in particular with the development of methods to hide such information. We argue that, when these texts are in fact messages exchanged between individuals, this is not the end of the story. Indeed, in this case, a second party, the intended recipient, is also involved and should be considered. In this work, we investigate the potential privacy leaks affecting them, that is we propose and address the problem of Recipient Profiling. We provide empirical evidence that such a task is feasible on several publicly accessible datasets (huggingface.co/datasets/sileod/recipient_profiling). Furthermore, we show that the learned models can be transferred to other datasets, albeit with a loss in accuracy.

8.2 Data Sets

A large-scale audit of dataset licensing and attribution in AI [18]

The race to train language models on vast, diverse and inconsistently documented datasets raises pressing legal and ethical concerns. To improve data transparency and understanding, we convene a multi-disciplinary effort between legal and machine learning experts to systematically audit and trace more than 1,800 text datasets. We develop tools and standards to trace the lineage of these datasets, including their source, creators, licences and subsequent use. Our landscape analysis highlights sharp divides in the composition and focus of data licenced for commercial use. Important categories including low-resource languages, creative tasks and new synthetic data all tend to be restrictively licenced. We observe frequent miscategorization of licences on popular dataset hosting sites, with licence omission rates of more than 70% and error rates of more than 50%. This highlights a crisis in misattribution and informed use of popular datasets driving many recent breakthroughs. Our analysis of data sources also explains the application of copyright law and fair use to finetuning data. As a contribution to continuing improvements in dataset transparency and responsible use, we release our audit, with an interactive user interface, the Data Provenance Explorer, to enable practitioners to trace and filter on data provenance for the most popular finetuning data collections: www.dataprovenance.org.

8.3 Decentralized Learning and security

SoK: Verifiable Cross-Silo FL [39]

Federated Learning (FL) is a widespread approach that allows training machine learning (ML) models with data distributed across multiple devices. 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 systematization of knowledge 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.

Verifiable cross-silo federated learning [43]

Federated Learning (FL) is a widespread approach that allows training machine learning (ML) models with data distributed across multiple devices. 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. However, malicious agents 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 conduct a comprehensive analysis of such protocols, and fit them in a taxonomy. We perform a comparison of the efficiency and threat models of various approaches. We next identify research gaps and discuss potential directions for future scientific work.

Protect-IT 2024 Workshop at the 37th IEEE Computer Security Foundations Symposium [43]

Federated Learning (FL) is a widespread approach that allows training machine learning (ML) models with data distributed across multiple devices. 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. However, malicious agents 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. This poster reflects the results decribed in the corresponding paper where we conduct a comprehensive analysis of such protocols, and fit them in a taxonomy. We also perform a comparison of the efficiency and threat models of various approaches. We next identify research gaps and discuss potential directions for future scientific work.

Overview of Secure Comparison [44]

Introduced by Yao's Millionaires' problem, Secure Comparison (SC) allows parties to compare two secrets in a privacy-preserving manner. This article gives an overview of the different SC techniques in various settings such as Secret Sharing (SS) or Homomorphic Encryption (HE).

Improved Stability and Generalization Guarantees of the Decentralized SGD Algorithm [20]

This paper presents a new generalization error analysis for Decentralized Stochastic Gradient Descent (D-SGD) based on algorithmic stability. The obtained results overhaul a series of recent works that suggested an increased instability due to decentralization and a detrimental impact of poorly-connected communication graphs on generalization. On the contrary, we show, for convex, strongly convex and non-convex functions, that D-SGD can always recover generalization bounds analogous to those of classical SGD, suggesting that the choice of graph does not matter. We then argue that this result is coming from a worst-case analysis, and we provide a refined optimization-dependent generalization bound for general convex functions. This new bound reveals that the choice of graph can in fact improve the worst-case bound in certain regimes, and that surprisingly, a poorly-connected graph can even be beneficial for generalization.

8.4 Privacy

Differential Privacy for Decentralized Learning [34]

The collapse of storage and data processing costs, along with the rise of digitization, has brought new applications and possibilities to machine learning. In practice, Big data is often synonymous with sensitive data collection. Hence, protecting privacy – especially by avoiding data leakage, intentional or accidental – is one of the key challenges in Trustworthy Machine Learning. A first direction towards more control over data is to keep it decentralized, exchanging only the information needed to run the learning process. This can be achieved through a central server orchestrating the learning process in federated learning or through peer-to-peer communications. However, this does not guarantee that data is protected throughout the entire process, as federated learning is known to be vulnerable to privacy attacks. To reliably quantify and control the privacy loss occurring in machine learning algorithms, Differential Privacy is currently the gold standard both in research and industry for machine learning applications. This thesis lies at the intersection of machine learning, decentralized algorithms and differential privacy. We present the first reconstruction attack in decentralized learning, targeting privacy leaks among participants not directly connected, proving the need to include defense mechanisms in this setting. We then introduce a new variant of differential privacy, Network Differential Privacy, which is suited for decentralized learning where each node only sees local communications. Using this variant, we analyze the privacy and utility guarantees of various decentralized algorithms, namely gossip algorithms and random walks for stochastic gradient descent, and ADMM. Our contributions demonstrate that decentralization can bring privacy amplification in the sense of differential privacy, and that the gains depend on the algorithm and the communication graph. This paves the way for the use of decentralization as a tool to develop more effective privacy-preserving machine learning.

Privacy-preserving AI using declarative constraints [33]

Machine learning and Deep learning-based technologies have gained widespread adoption, quickly displacing traditional artificially intelligent (AI) systems. Contemporary computers are remarkable in processing enormous amounts of personal data through these machine learning (ML) algorithms. However, this technological advancement brings along significant privacy implications, and this problem can only be expected to escalate in the foreseeable future.Studies have shown that it is possible to deduce sensitive information from statistical models computed on datasets, even without direct access to the underlying training dataset. Apart from the privacy-related concerns regarding statistical models, the complex systems learning and employing such models are increasingly difficult for users to understand, and so are the ramifications of consenting to the submission and use of their private information within such frameworks. Consequently, transparency and interpretability emerged as pressing concerns.In this dissertation, we study the problem of specifying privacy requirements for machine learning based systems, in a manner that combines interpretability with operational feasibility. Explaining privacy-improving technology is a challenging problem, especially when the objective is to construct a system that at the same time is interpretable and has a high utility. In order to address this challenge, we propose to specify privacy requirements as constraints, thereby allowing for both interpretability and automated optimization of the utility.

A Smartphone-based Architecture for Prolonged Monitoring of Gait [30]

Gait analysis is important for evaluating neurological disorders such as stroke and Parkinson's disease. Traditionally, healthcare professionals had to rely on subjective assessments (i.e., human-based) of gait which were time consuming and not very reproducible. However, with the advent of IoT and indeed more objective (e.g., measurement-based) assessment methods, gait analysis can now be performed more accurately and effectively. It is worth noting, however, that there are still limitations to these objective methods, especially the lack of privacy-preserving continuous data collection. To overcome this limitation, we present in this paper a privacy-by-design monitoring application for post-stroke patients to evaluate their gait before, during, and after a rehabilitation program. Gait measurements are collected by a mobile application that continuously captures spatiotemporal parameters in the background using the built-in smartphone accelerometer. Statistical techniques are then applied to extract general indicators about the performed activity, as well as some more specific gait metrics in real-time such as regularity, symmetry and walking speed. These metrics are calculated based on the detected steps while patients are performing an activity. Additionally, a deep learning approach based on an auto-encoder is implemented to detect abnormal activities in the gait of patients. These analyses provides both valuable insights and statistical information about the activities performed by the patient, and a useful tool for practitioners to monitor the progression of neurological disorders and detect anomalies. We conducted experiments using this application in real conditions to monitor post-stroke patients in collaboration with a hospital, demonstrating its ability to compute valuable metrics and detect abnormal events patient's gait.

Honest Fraction Differential Privacy [28]

Over the last decades, differential privacy (DP) has become a standard notion of privacy. It allows to measure how much sensitive information an adversary could infer from a result (statistical model, prediction, etc.) he obtains. In privacy-preserving federated machine learning, one aims to learn a statistical model from data owned by multiple data owners without revealing their sensitive data. A common strategy is to use secure multi-party computation (SMPC) to avoid revealing intermediate results. However, DP assumes a very strong adversary who is able to know all information in the dataset except the targeted secret, while most SMPC methods assume a clearly less strong adversary, e.g., it is common to assume that the adversary has bounded computational power and can corrupt only a minority of the data owners (honest majority). As a chain is not stronger than its weakest part, in such combinations the DP provides an overly strong protection at an unnecessarily high cost in terms of utility. We propose honest fraction differential privacy, which is similar to differential privacy but assumes that the adversary can only collude with data owners covering part of the data. This assumption is very similar to the assumptions made by many SMPC strategies. We illustrate this idea by considering the application to the specific task of unregularized linear regression without bias on sufficiently large datasets.

Rényi Pufferfish Privacy: General Additive Noise Mechanisms and Privacy Amplification by Iteration via Shift Reduction Lemmas [26]

Pufferfish privacy is a flexible generalization of differential privacy that allows to model arbitrary secrets and adversary's prior knowledge about the data. Unfortunately, designing general and tractable Pufferfish mechanisms that do not compromise utility is challenging. Furthermore, this framework does not provide the composition guarantees needed for a direct use in iterative machinelearning algorithms. To mitigate these issues, we introduce a Rényi divergence-based variant of Pufferfish and show that it allows us to extend the applicability of the Pufferfish framework. We first generalize the Wasserstein mechanism to cover a wide range of noise distributions and introduce several ways to improve its utility. Finally, as an alternative to composition, we prove privacy amplification results for contractive noisy iterations and showcase the first use of Pufferfish in private convex optimization. A common ingredient underlying our results is the use and extension of shift reduction lemmas.

Privacy Attacks in Decentralized Learning [24]

Decentralized Gradient Descent (D-GD) allows a set of users to perform collaborative learning without sharing their data by iteratively averaging local model updates with their neighbors in a network graph.

The absence of direct communication between non-neighbor nodes might lead to the belief that users cannot infer precise information about the data of others. In this work, we demonstrate the opposite, by proposing the first attack against D-GD that enables a user (or set of users) to reconstruct the private data of other users outside their immediate neighborhood. Our approach is based on a reconstruction attack against the gossip averaging protocol, which we then extend to handle the additional challenges raised by D-GD. We validate the effectiveness of our attack on real graphs and datasets, showing that the number of users compromised by a single or a handful of attackers is often surprisingly large. We empirically investigate some of the factors that affect the performance of the attack, namely the graph topology, the number of attackers, and their position in the graph.

DP-SGD with weight clipping

Recently, due to the popularity of deep neural networks and other methods whose training typically relies on the optimization of an objective function, and due to concerns for data privacy, there is a lot of interest in differentially private gradient descent methods. To achieve differential privacy guarantees with a minimum amount of noise, it is important to be able to bound precisely the sensitivity of the information which the participants will observe. In this study, we present a novel approach that mitigates the bias arising from traditional gradient clipping. By leveraging a public upper bound of the Lipschitz value of the current model and its current location within the search domain, we can achieve refined noise level adjustments. We present a new algorithm with improved differential privacy guarantees and a systematic empirical evaluation, showing that our new approach outperforms existing approaches also in practice.

Optimal Classification under Performative Distribution Shift [31]

Performative learning addresses the increasingly pervasive situations in which algorithmic decisions may induce changes in the data distribution as a consequence of their public deployment. We propose a novel view in which these performative effects are modelled as push-forward measures. This general framework encompasses existing models and enables novel performative gradient estimation methods, leading to more efficient and scalable learning strategies. For distribution shifts, unlike previous models which require full specification of the data distribution, we only assume knowledge of the shift operator that represents the performative changes. This approach can also be integrated into various change-of-variablebased models, such as VAEs or normalizing flows. Focusing on classification with a linear-in-parameters performative effect, we prove the convexity of the performative risk under a new set of assumptions. Notably, we do not limit the strength of performative effects but rather their direction, requiring only that classification becomes harder when deploying more accurate models. In this case, we also establish a connection with adversarially robust classification by reformulating the minimization of the performative risk as a min-max variational problem. Finally, we illustrate our approach on synthetic and real datasets.

Confidential-DPproof: Confidential Proof of Differentially Private Training [27]

Post hoc privacy auditing techniques can be used to test the privacy guarantees of a model, but come with several limitations: (i) they can only establish lower bounds on the privacy loss, (ii) the intermediate model updates and some data must beshared with the auditor to get a better approximation of the privacy loss, and (iii) the auditor typically faces a steep computational cost to run a large number of attacks. In this paper, we propose to proactively generate a cryptographic certificate of privacy during training to forego such auditing limitations. We introduce Confidential-DPproof , a framework for Confidential Proof of Differentially Private Training, which enhances training with a certificate of the (ϵ, δ)-DP guarantee achieved. To obtain this certificate without revealing information about the training data or model, we design a customized zero-knowledge proof protocol tailored to the requirements introduced by differentially private training, including random noise addition and privacy amplification by subsampling. In experiments on CIFAR-10, Confidential-DPproof trains a model achieving state-of-the-art 91% test accuracy with a certified privacy guarantee of (ϵ = 0.55, δ = 10–5)-DP in approximately 100 hours.

Differentially Private Decentralized Learning with Random Walks [19]

The popularity of federated learning comes from the possibility of better scalability and the ability for participants to keep control of their data, improving data security and sovereignty. Unfortunately, sharing model updates also creates a new privacy attack surface. In this work, we characterize the privacy guarantees of decentralized learning with random walk algorithms, where a model is updated by traveling from one node to another along the edges of a communication graph. Using a recent variant of differential privacy tailored to the study of decentralized algorithms, namely Pairwise Network Differential Privacy, we derive closed-form expressions for the privacy loss between each pair of nodes where the impact of the communication topology is captured by graph theoretic quantities. Our results further reveal that random walk algorithms tends to yield better privacy guarantees than gossip algorithms for nodes close from each other. We supplement our theoretical results with empirical evaluation on synthetic and real-world graphs and datasets.

Analysis of Speech Temporal Dynamics in the Context of Speaker Verification and Voice Anonymization [29]

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.

8.5 Fairness and Transparency

On the Impact of Output Perturbation on Fairness in Binary Linear Classification [38]

We theoretically study how differential privacy interacts with both individual and group fairness in binary linear classification. More precisely, we focus on the output perturbation mechanism, a classic approach in privacy-preserving machine learning. We derive high-probability bounds on the level of individual and group fairness that the perturbed models can achieve compared to the original model. Hence, for individual fairness, we prove that the impact of output perturbation on the level of fairness is bounded but grows with the dimension of the model. For group fairness, we show that this impact is determined by the distribution of so-called angular margins, that is signed margins of the non-private model re-scaled by the norm of each example.

Measuring and Mitigating Allocation Unfairness Across the Machine Learning Pipeline [35]

With the advent of machine learning, the government institutions and other bureaucracy are undergoing a paradigm shift, as algorithms increasingly assist in and even replace some of their functions. Consequently, just as early 20th-century philosophers scrutinized these institutional changes, it is crucial to analyze these algorithms through the lens of their societal impact.In line with this general objective, this thesis aims to examine and propose ways to mitigate the harms associated with employing machine learning (ML). Specifically, we study the impact of ML algorithm in the settings where groups of population are unfairly assigned or withheld opportunities and resources. In response, we propose a series of algorithms designed to measure and counteract unfairness throughout the ML pipeline. We begin by proposing FairGrad, a gradient based algorithm which dynamically adjusts the influence of examples throughout the training process to ensure fairness. We then examine FairGrad, and various other fairness enforcing mechanism from the lens of intersectionality where multiple sensitive demographic attributes are considered together. Our experiments reveal that several approaches exhibit "leveling down" behavior, implying that they optimize for current fairness measures by harming the involved groups. We introduce a new fairness measure called [dollar]alpha[dollar]-Intersectional Fairness which helps uncover this phenomena.Building upon these findings, our next step focuses on addressing the leveling down issue. To mitigate its effects, we introduce a data generation mechanism that exploits the hierarchial structure

inherent to the intersectional setting, and augments data for groups by combining and transforming data from more general groups. Through our experiments we find that this approach not only produces realistic new examples but also enhances performance in worst-case scenarios. Finally, we explore the intersection of privacy, another societal concern, with fairness. We present FEDERATE, a novel method that combines adversarial learning with differential privacy to derive private representations that lead to fairer outcomes. Interestingly, our results suggest that in our experimental context privacy and fairness can coexist and frequently complement each other.

Synthetic Data Generation for Intersectional Fairness by Leveraging Hierarchical Group Structure [41]

In this paper, we introduce a data augmentation approach specifically tailored to enhance intersectional fairness in classification tasks. Our method capitalizes on the hierarchical structure inherent to intersectionality, by viewing groups as intersections of their parent categories. This perspective allows us to augment data for smaller groups by learning a transformation function that combines data from these parent groups. Our empirical analysis, conducted on four diverse datasets including both text and images, reveals that classifiers trained with this data augmentation approach achieve superior intersectional fairness and are more robust to "leveling down" when compared to methods optimizing traditional group fairness metrics.

8.6 Machine Learning

Central Limit Theorem for Bayesian Neural Network trained with Variational Inference [37]

In this paper, we rigorously derive Central Limit Theorems (CLT) for Bayesian two-layerneural networks in the infinite-width limit and trained by variational inference on a regression task. The different networks are trained via different maximization schemes of the regularized evidence lower bound: (i) the idealized case with exact estimation of a multiple Gaussian integral from the reparametrization trick, (ii) a minibatch scheme using Monte Carlo sampling, commonly known as Bayes-by-Backprop, and (iii) a computationally cheaper algorithm named Minimal VI. The latter was recently introduced by leveraging the information obtained at the level of the mean-field limit. Laws of large numbers are already rigorously proven for the three schemes that admits the same asymptotic limit. By deriving CLT, this work shows that the idealized and Bayes-by-Backprop schemes have similar fluctuation behavior, that is different from the Minimal VI one. Numerical experiments then illustrate that the Minimal VI scheme is still more efficient, in spite of bigger variances, thanks to its important gain in computational complexity.

8.7 Theoretical Computer Science

Linear Programs with Conjunctive Database Queries [17]

In this paper, we study the problem of optimizing a linear program whose variables are the answers to a conjunctive query. For this we propose the language LP(CQ) for specifying linear programs whose constraints and objective functions depend on the answer sets of conjunctive queries. We contribute an efficient algorithm for solving programs in a fragment of LP(CQ). The natural approach constructs a linear program having as many variables as there are elements in the answer set of the queries. Our approach constructs a linear program having the same optimal value but fewer variables. This is done by exploiting the structure of the conjunctive queries using generalized hypertree decompositions of small width to factorize elements of the answer set together. We illustrate the various applications of LP(CQ) programs on three examples: optimizing deliveries of resources, minimizing noise for differential privacy, and computing the s-measure of patterns in graphs as needed for data mining.

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 Siléo, 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 European initiatives

10.1.1 Horizon Europe

TRUMPET TRUMPET project on cordis.europa.eu

Title: TRUstworthy Multi-site Privacy Enhancing Technologies

Duration: From October 1, 2022 to September 30, 2025

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- TIMELEX (TIMELEX), Belgium
- TECHNOVATIVE SOLUTIONS LTD, United Kingdom
- FUNDACION CENTRO TECNOLOXICO DE TELECOMUNICACIONS DE GALICIA (GRADIANT), Spain
- COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES (CEA), France
- ISTITUTO ROMAGNOLO PER LO STUDIO DEI TUMORI DINO AMADORI - IRST SRL (IRST), Italy
- CENTRE HOSPITALIER UNIVERSITAIRE DE LIEGE (CHUL), Belgium
- Turkiye Cumhuriyeti Saglik Bakanligi (MOH), Türkiye
- UNIVERSIDAD DE VIGO (UVIGO), Spain
- ARTEEVO TECHNOLOGIES LTD (ARTEEVO), Israel

Inria contact: Jan Ramon

Coordinator: Gradiant

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 FLUTE project on cordis.europa.eu

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

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

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- QUIBIM SOCIEDAD LIMITADA (QUIBIM), Spain
- TIMELEX (TIMELEX), Belgium
- TECHNOVATIVE SOLUTIONS LTD, United Kingdom
- HL7 INTERNATIONAL FONDATION (HL7 INTERNATIONAL), Belgium
- FUNDACION CENTRO TECNOLOXICO DE TELECOMUNICACIONS DE GALICIA (GRADIANT), Spain
- SIEMENS SRL, Romania
- UNIVERSITAT POLITECNICA 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: Jan Ramon (INRIA)

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.

10.2 National initiatives

10.2.1 ANR PMR (2020-2024)

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

10.2.2 FedMalin. INRIA Defi (2021-2024)

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.

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

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

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).

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

Participants: Jan Ramon, Marc Tommasi *(contact person)*, Michaël Perrot, Cesar Sabater, Edwige Cyffers, 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.

10.2.5 CAPS'UL (2023-2028)

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

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.

10.2.6 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.

10.2.7 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.

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.

10.2.8 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 userdefined 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.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

- MICHAËL PERROT, MIKAELA KELLER and MARC TOMMASI have organized CAp-RFIAP’2024.
- PASCAL DENIS has co-organized the 1st Annual LLcD (Language and Languages at the crossroads of Disciplines) Meeting.

11.1.2 Scientific events: selection

- MARC TOMMASI served as Area Chair for UAI, PC member of AISTATS and APVP.
- MICHAËL PERROT served as Reviewer for ICML 2024, AAAI 2024.

- JAN RAMON was PC member of AFME@NeurIPS 2024, AISTATS 2024, BNAIC 2024, ECML/PKDD 2024, ICLR 2024, ICML 2024, IJCAI 2024, IJCAI-DC 2024, MLG@ECML 2024, NeurIPS 2024, ProtectIT@CSF 2024, SDM 2024, UAI 2024.
- MIKAELA KELLER served as Reviewer for ACL Rolling Review.
- DAMIEN SILEO served as Reviewer for ACL Rolling Review.
- PASCAL DENIS was Action Editor for ACL Rolling Review, and served as PC member for COLING-LREC 2024, EMNLP 2024, NAACL 2024.
- BATISTE LE BARS served as Reviewer for AISTATS 2024.

11.1.3 Journal

- JAN RAMON is member of the editorial boards of Machine Learning Journal (MLJ), Data Mining and Knowledge Discovery (DMKD), ECML-PKDD Journal track.

Reviewer - reviewing activities

- JAN RAMON is reviewer of Transactions of Machine Learning Research (TMLR) and reviews individual submissions for other journals.
- PASCAL DENIS is standing reviewer for Transactions of the Association for Computational Linguistics (TACL).
- BATISTE LE BARS served as a reviewer for the Journal of Machine Learning Research (JMLR) and for the SIAM Journal on Optimization.

11.1.4 Invited talks

- DAMIEN SILEO gave a seminar at CRIL (Lens).
- MARC TOMMASI gave a seminar on responsible AI in a colloquium on "Droit et éthique": in the SCAI colloquium (Paris); in the INRIA-Brasil meeting (online); in a local workshop on Neuroscience and AI.
- BATISTE LE BARS gave a seminar on Federated Conformal Prediction at Owkin and a seminar on the generalization properties of decentralized algorithms at the Laboratoire de Mathématiques d'Orsay.
- JAN RAMON: 'Applying differential privacy theory in practical applications', Protect-IT workshop at CSF-2024, Enschede, NL.
- JAN RAMON presented the FLUTE project he coordinates and related topics at, among others,
  - RERI Numerique (4/11/2024)
  - BVDA's trustworthy AI task force ETAMI (8/5/2024)
  - EC HADEA's workshop for projects in the Health programme (25/4/2024)

11.1.5 Leadership within the scientific community

- JAN RAMON was member of the bureau of the Societé Savante Francophone d'Apprentissage Machine (SSFAM)

11.1.6 Scientific expertise

- MARC TOMMASI was a member (scientific expert) of the recruitment committee of assistant professors at Saint-Etienne.
- MIKAELA KELLER was a member (scientific expert) of the recruitment committee of assistant professors at Saint-Etienne and Lille Universities and the recruitement commitee of CRCN and ISFP in Inria Lille.
- JAN RAMON was reviewer for COST project proposals and Horizon Europe project monitoring.

11.1.7 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 co-head of the CNRS GDR "Langues et langage à la croisée des disciplines" (LLcD) and a member of the CNRS GDR NLP Group. PASCAL DENIS is also a member of the network "référents données" at Inria and Université de Lille (Lille Open Research Data). He is administrator of Inria membership to Linguistic Data Consortium (LDC).
- MICHAËL PERROT is a substitue member of the Comité de Centre (named, representing the administration). MICHAËL PERROT is volunteer in the local AGOS team.

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- Licence MIASHS: MARC TOMMASI, Data Science, 24h, L2, Université de Lille.
- Licence MIASHS: MIKAELA KELLER, Fouille de graphes, 24h, L3, 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 language naturel, 24h, M2, Université de Lille.
- Master Computer Science: MARC TOMMASI, Data Science, 48h, M1, Université de Lille.
- Master Computer Science: MARC TOMMASI, Semi-supervised learning and Graphs, 24h, M2, 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.
- MARC TOMMASI is directeur des études for the Machine Learning master of Computer Science.

11.2.2 Supervision

- Postdoc: VITALII EMILIANOV. On the interactions between fairness and privacy in machine learning. Nov. 22-Jun 24. MICHAËL PERROT.
- Postdoc: ARNAUD DESCOURS. On federated optimization with lower communication cost. Since Nov. 2023. JAN RAMON.
- Postdoc: IMANE TALBI. On the relation between statistical privacy and security assumptions. May 23-Apr. 24. JAN RAMON.
- Postdoc: LUIS LUGO. On federated learning with energy budgets, since Jun. 24. MARC TOMMASI and ROMAIN ROUVOY.
- PhD defended in Apr. 24: MOITREE BASU, Integrated privacy-preserving AI, since 2019. JAN RAMON [33].
- Phd defended in Mar. 24: GAURAV MAHESHWARI. Trustworthy Representations for Natural Language Processing, since Nov 2020. AURÉLIEN BELLET, MIKAELA KELLER, and PASCAL DENIS [35].
- Phd defended in Dec. 24: EDWIGE CYFFERS. Decentralized learning and privacy amplification, since Oct. 2021. AURÉLIEN BELLET [34].
- Phd in progress: SHREYA VENUGOPAL. Guaranteed Fairness in Machine Learning, since Oct. 24, MICHAËL PERROT.
- Phd in progress: JEAN DUFRAICHE. Fairness and Privacy in Machine Learning, since Oct. 24, MICHAËL PERROT, MARC TOMMASI.
- Phd in progress: MARC DAMIE. Secure protocols for verifiable decentralized machine learning, since May 2022. JAN RAMON with Andreas Peter (U. Twente, NL & U. Oldenburg, DE), Florian Hahn (University of Twente, NL).
- 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: DINH-VIET-TOAN LE. Natural Language Processing approaches in the musical domain : suitability, performance and limits, since Oct. 2022. MIKAELA KELLER and LOUIS BIGO.
- Phd in progress: ALEKSEI KORNEEV. Trustworthy multi-site privacy-enhancingtechnologies, since Dec. 2022. JAN RAMON.
- PhD in progress: ANTOINE BARCZEWSKI. Transparent privacy-preserving machine learning, since May 2022. JAN RAMON.
- PhD in progress: AURÉLIEN SAÏD HOUSSEINI. Computational Models of Semantic Memory, since Sept. 2023. ANGÈLE BRUNELLIÈRE (UMR SCALab, Université de Lille), PASCAL DENIS and RÉMI GILLERON.
- 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 anonymizaton, 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.
- Engineer JULES BOULET, FLUTE and TRUMPET projects, since Jun. 24, 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 ELINA THIBEAU-SUTRE, FLUTE and TRUMPET projects, since May. 24, JAN RAMON.
- Engineer SOPHIE VILLEROT, ADT project Tailed: Trustworthy AI Library for Environments which are Decentralized, Nov. 2020-Apr. 24., JAN RAMON.
- Engineer PAUL ANDREY, Decentralized and Federated Learning with DecLearn. Jul. 2022-Oc 24. AURÉLIEN BELLET and MARC TOMMASI. Now Phd since Nov. 24, MARC TOMMASI.
- Engineer QUENTIN SINH. Integrating MAGNET results in the TRUMPET privacy-preserving federated learning platform, Since Nov 2023. JAN RAMON. Now Phd since Jul 24, JAN RAMON.
- Engineer KEVIN NGAKOSSO. Integrating MAGNET results on constraint-based privacy in the TRUMPET privacy-preserving federated learning platform, Oct. 2023-Nov. 24, JAN RAMON.
- Engineer LÉONARD DEROOSE. Development of TRUMPET privacy-preserving platform components, Since Sep. 2023. JAN RAMON.
- Engineer LI MOU. Statistical privacy computation algorithms, Feb. 2023-May-24, JAN RAMON.

11.2.3 Juries

- MARC TOMMASI member of the PhD jury of Patrick Saux (President), Paul Mangold (Examiner), Philippo Galli (Reviewer), Konstandinos Aiwansedo (Reviewer).
- MICHAËL PERROT member of the PhD jury of Karima Makhlouf (Examiner), Jan Aalmoes (Examiner).
- PASCAL DENIS was member of the PhD committee (Examiner) of Alban Petit, LISN, Université Paris-Saclay

11.3 Popularization

11.3.1 Participation in Live events

- JAN RAMON gave a presentation at “Bar des Sciences” at the University of Lille on ’Apprentisage fédéré des données internationaux avec applications au cancer de prostate’, touching societal questions related to trustworthy AI.

11.3.2 Others science outreach relevant activities

- MICHAËL PERROT gave an Introduction to Machine Learning during a Journée du Numérique on Intelligence artificielle : enjeux éducatifs et pistes pédagogiques organized by INSPÉ Lille HdF (Recording).
- MICHAËL PERROT and DAMIEN SILEO participated in a round table during a Journée du Numérique on Intelligence artificielle : enjeux éducatifs et pistes pédagogiques organized by INSPÉ Lille HdF.
- MIKAELA KELLER is an organizer of Pixelles, a coding and computer science workshop for teenage girls taking place every week since October.

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. 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. Kużelka, 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] 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.

[9] C. Pelekis, J. Ramon and Y. Wang. 'Hˊolder-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. URL: https://hal.archives-ouvertes.fr/hal-01421953.

[10] T. Ricatte, R. Gilleron and M. Tommasi. 'Skill Rating for Multiplayer Games Introducing Hypernode Graphs and their Spectral Theory'. In: *Journal of Machine Learning Research* 21 (2020), pp. 1–18. URL: https://hal.inria.fr/hal-02566930.

[11] 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. URL: https://hal.inria.fr/hal-03820603.

[12] 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.

[13] 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.

[14] 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.

[15] F. Vitale, N. Parotsidis and C. Gentile. 'Online Reciprocal Recommendation with Theoretical Performance Guarantees'. In: *NIPS 2018 - 32nd Conference on Neural Information Processing Systems*. Montreal, Canada, Dec. 2018. URL: https://hal.inria.fr/hal-01916979.

12.2 Publications of the year

International journals

[16] L. Bigo, M. Keller and D.-V.-T. Le. 'Le langage des partitions musicales face à l'intelligence artificielle'. In: *Culture et recherche* 147 (21st Nov. 2024), pp. 34–35. URL: https://hal.science/hal-04909478.

[17] F. Capelli, N. Crosetti, J. Niehren and J. Ramon. 'Linear Programs with Conjunctive Database Queries'. In: *Logical Methods in Computer Science* Volume 20, Issue 1 (3rd Jan. 2024). DOI: 10.46298/lmcs-20(1:9)2024. URL: https://hal.science/hal-04317553. In press (cit. on p. 19).

[18] 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. URL: https://hal.science/hal-04749695 (cit. on pp. 10, 13).

International peer-reviewed conferences

[19] E. Cyffers, A. Bellet and J. Upadhyay. 'Differentially Private Decentralized Learning with Random Walks'. In: ICML 2024 - Forty-first International Conference on Machine Learning. Vienne (Autriche), Austria: arXiv, 2024. DOI: 10.48550/arXiv.2402.07471. URL: https://hal.science/hal-04610660 (cit. on p. 18).

[20] 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 (cit. on p. 15).

[21] B. Liétard, P. Denis and M. Keller. 'To Word Senses and Beyond: Inducing Concepts with Contextualized Language Models'. In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2024 - Conference on Empirical Methods in Natural Language Processing. Miami, United States: Association for Computational Linguistics, 12th Nov. 2024, pp. 2684–2696. DOI: 10.18653/v1/2024.emnlp-main.156. URL: https://inria.hal.science/hal-04848884 (cit. on p. 13).

[22] B. Liétard, M. Keller and P. Denis. 'Towards an Onomasiological Study of Lexical Semantic Change through the Induction of Concepts'. In: *Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change*. 5th Workshop on Computational Approaches to Historical Language Change. Bangkok, Thailand, Thailand: Association for Computational Linguistics; Association for Computational Linguistics, Aug. 2024, pp. 158–167. DOI: 10.18653/v1/2024.lchange-1.15. URL: https://inria.hal.science/hal-04681251 (cit. on p. 12).

[23] S. Longpre, R. Mahari, A. Lee, C. Lund, H. Oderinwale, W. Brannon, N. Saxena, N. Obeng-Marnu, T. South, C. Hunter, K. Klyman, C. Klamm, H. Schoelkopf, N. Singh, M. Cherep, A. Anis, A. Dinh, C. Chitongo, D. Yin, D. Sileo, D. Mataciunas, D. Misra, E. Alghamdi, E. Shippole, J. Zhang, J. Materzynska, K. Qian, K. Tiwary, L. Miranda, M. Dey, M. Liang, M. Hamdy, N. Muennighoff, S. Ye, S. Kim, S. Mohanty, V. Gupta, V. Sharma, V. M. Chien, X. Zhou, Y. Li, C. Xiong, L. Villa, S. Biderman, H. Li, D. Ippolito, S. Hooker, J. Kabbara and S. Pentland. 'Consent in Crisis: The Rapid Decline of the AI Data Commons'. In: NEURIPS 2024 - Thirty-Eighth Annual Conference on Neural Information Processing Systems. Vancouver (CA), Canada, 2024. URL: https://hal.science/hal-04824161 (cit. on pp. 10, 11).

[24] A. E. Mrini, E. Cyffers and A. Bellet. 'Privacy Attacks in Decentralized Learning'. In: ICML 2024 - Forty-first International Conference on Machine Learning. Vienne (Austria), Austria: arXiv, 2024. DOI: 10.48550/arXiv.2402.10001. URL: https://hal.science/hal-04610652 (cit. on p. 16).

[25] C. Oguz, P. Denis, S. Ostermann, N. Skachkova, E. Vincent and J. van Genabith. 'MMAR: Multilingual and multimodal anaphora resolution in instructional videos'. In: Findings of the 2024 Conference on Empirical Methods in Natural Language Processing. Miami, United States, 12th Nov. 2024. URL: https://inria.hal.science/hal-04733760 (cit. on p. 12).

[26] C. Pierquin, A. Bellet, M. Tommasi and M. Boussard. 'Rényi Pufferfish Privacy: General Additive Noise Mechanisms and Privacy Amplification by Iteration via Shift Reduction Lemmas'. In: International Conference on Machine Learning (ICML 2024). Vienna (Austria), Austria, 2024. URL: https://inria.hal.science/hal-04363020 (cit. on p. 16).

[27] A. S. Shamsabadi, G. Tan, T. I. Cebere, A. Bellet, H. Haddadi, N. Papernot, X. Wang and A. Weller. 'Confidential-DPproof: Confidential Proof of Differentially Private Training'. In: ICLR 2024 - 12th International Conference on Learning Representations. Vienna (Austria), Austria, 2024. URL: https://hal.science/hal-04610635 (cit. on p. 17).

[28] I. Taibi and J. Ramon. 'Honest Fraction Differential Privacy'. In: *Proceedings of the 2024 ACM workshop on information hiding and multimedia security*. ACM workshop on information hiding and multimedia security. Vigo, Spain, 12th June 2024, pp. 1–5. DOI: 10.1145/3658664.3659655. URL: https://inria.hal.science/hal-04610199 (cit. on p. 16).

[29] 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. URL: https://hal.science/hal-04853872 (cit. on p. 18).

Conferences without proceedings

[30] L. Bart, E. A. Bechorfa, A. Boutet, J. Ramon and C. Frindel. 'A Smartphone-based Architecture for Prolonged Monitoring of Gait'. In: 2024 IEEE First International Conference on Artificial Intelligence for Medicine, Health and Care (AIMHC). Laguna Hills, United States: IEEE, 5th Feb. 2024, pp. 16–17. DOI: 10.1109/AIMHC59811.2024.00010. URL: https://hal.science/hal-04909717 (cit. on p. 16).

[31] E. Cyffers, M. S. Pydi, J. Atif and O. Cappé. 'Optimal Classification under Performative Distribution Shift'. In: 38th Conference on Neural Information Processing Systems. Vancouver (Canada), Canada, 10th Dec. 2024. URL: https://hal.science/hal-04763797 (cit. on p. 17).

[32] D.-V.-T. Le, L. Bigo and M. Keller. 'Analyzing Byte-Pair Encoding on Monophonic and Polyphonic Symbolic Music: A Focus on Musical Phrase Segmentation'. In: 3rd Workshop on NLP for Music and Audio (NLP4MusA). San Francisco, United States, 15th Nov. 2024. URL: https://hal.science/hal-04710532 (cit. on p. 11).

Doctoral dissertations and habilitation theses

[33] M. Basu. 'Privacy-preserving AI using declarative constraints'. Université de Lille, 4th Apr. 2024. URL: https://hal.science/tel-04621995 (cit. on pp. 15, 28).

[34] E. Cyffers. 'Differential Privacy for Decentralized Learning'. Université de Lille 1, Sciences et Technologies; CRIStAL UMR 9189, 5th Dec. 2024. URL: https://hal.science/tel-04907994 (cit. on pp. 15, 28).

[35] G. Maheshwari. 'Measuring and Mitigating Allocation Unfairness Across the Machine Learning Pipeline'. Université de Lille, 27th Mar. 2024. URL: https://hal.science/tel-04623248 (cit. on pp. 18, 28).

Reports & preprints

[36] M. Borquez, M. Keller, M. Perrot and D. Sileo. *Recipient Profiling: Predicting Characteristics from Messages*. 17th Dec. 2024. URL: https://hal.science/hal-04850698 (cit. on p. 13).

[37] A. Descours, T. Huix, A. Guillin, M. Michel, É. Moulines and B. Nectoux. *Central Limit Theorem for Bayesian Neural Network trained with Variational Inference*. 3rd June 2024. URL: https://hal.science/hal-04599502 (cit. on p. 19).

[38] V. Emelianov and M. Perrot. *On the Impact of Output Perturbation on Fairness in Binary Linear Classification*. 5th Feb. 2024. URL: https://inria.hal.science/hal-04440982 (cit. on p. 18).

[39] A. Korneev and J. Ramon. *SoK: Verifiable Cross-Silo FL*. 9th Sept. 2024. URL: https://hal.science/hal-04692331 (cit. on p. 14).

[40] D.-V.-T. Le, L. Bigo, M. Keller and D. Herremans. *Natural Language Processing Methods for Symbolic Music Generation and Information Retrieval: a Survey*. 27th Feb. 2024. URL: https://hal.science/hal-04621444 (cit. on p. 12).

[41] G. Maheshwari, A. Bellet, P. Denis and M. Keller. *Synthetic Data Generation for Intersectional Fairness by Leveraging Hierarchical Group Structure*. 23rd May 2024. URL: https://inria.hal.science/hal-04863199 (cit. on p. 19).

[42] D. Sileo. *Scaling Synthetic Logical Reasoning Datasets with Context-Sensitive Declarative Grammars*. 20th June 2024. URL: https://hal.science/hal-04618305 (cit. on p. 12).

Other scientific publications

[43] A. Korneev and J. Ramon. 'Verifiable cross-silo federated learning'. In: Protect-IT 2024 Workshop at the 37th IEEE Computer Security Foundations Symposium. Enschede (Pays Bas), France, 8th July 2024. URL: https://hal.science/hal-04612742 (cit. on p. 14).

[44] Q. Sinh and J. Ramon. *Overview of Secure Comparison*. 14th June 2024. URL: https://hal.science/hal-04612505 (cit. on p. 14).