RESEARCH CENTRE

Inria Centre at the University of Lille

IN PARTNERSHIP WITH:

CNRS, Université de Lille

2023

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

- Aurélien Bellet [INRIA, Researcher, until Jul 2023, HDR]
- Pascal Denis [INRIA, Researcher]
- Michaël Perrot [INRIA, ISFP]
- Jan Ramon [INRIA, Senior Researcher, HDR]
- Damien Sileo [INRIA, ISFP]

Faculty Members

- Marc Tommasi [Team leader, UNIV LILLE, Professor Delegation, HDR]
- Mikaela Keller [UNIV LILLE, Associate Professor Delegation, from Sep 2023]
- Mikaela Keller [UNIV LILLE, Associate Professor, until Aug 2023]

Post-Doctoral Fellows

- Sabri Chellouf [INRIA, Post-Doctoral Fellow, from May 2023]
- Baptiste Cottier [INRIA, Post-Doctoral Fellow, from May 2023 until Sep 2023]
- Arnaud Descours [INRIA, Post-Doctoral Fellow, from Nov 2023]
- Vitalii Emelianov [INRIA, Post-Doctoral Fellow]
- Batiste Le Bars [INRIA, Post-Doctoral Fellow, until May 2023]
- Cesar Sabater [INRIA, Post-Doctoral Fellow, until Mar 2023]
- Imane Taibi [INRIA, Post-Doctoral Fellow, from May 2023]

PhD Students

- Antoine Barczewski [UNIV LILLE]
- Moitree Basu [INRIA]
- Ioan Tudor Cebere [INRIA]
- Edwige Cyffers [UNIV LILLE]
- Marc Damie [Univ. of Twente, from Oct 2023]
- Marc Damie [INRIA, until Sep 2023]
- Le Dinh-Viet-Toan [UNIV LILLE]
- Brahim Erraji [INRIA, from Sep 2023]
- Aleksei Korneev [UNIV LILLE]
- Bastien Lietard [INRIA]
- Gabriel Loiseau [Vade, CIFRE, from Sep 2023]
- Gaurav Maheshwari [INRIA]

- Paul Mangold [INRIA, until Oct 2023]
- Amal Mawass [UNIV LILLE]
- Clément Pierquin [CRAFT, CIFRE, from Jun 2023]
- Arijus Pleska [INRIA, until Apr 2023]
- Aurélien Said Housseini [INRIA, from Sep 2023]

Technical Staff

- Paul Andrey [INRIA, Engineer]
- Nathan Bigaud [INRIA, Engineer, until Aug 2023]
- Léonard Deroose [INRIA, Engineer, from Sep 2023]
- Rishabh Gupta [INRIA, Engineer, until Jun 2023]
- Mou Li [INRIA, Engineer, from Feb 2023]
- Kamalkumar Ramanlal Macwan [UNIV LILLE, Engineer, until Sep 2023]
- Kevin Hubert N Gakosso [INRIA, Engineer, from Oct 2023]
- Joseph Renner [INRIA, Engineer, until Sep 2023]
- Quentin Sinh [INRIA, Engineer, from Nov 2023]
- Sophie Villerot [INRIA, Engineer]

Interns and Apprentices

- Mathis Allard-Mevel [INRIA, Intern, from Apr 2023 until Sep 2023]
- Nilia Bardachene [INRIA, Intern, from Apr 2023 until Aug 2023]
- Louise Bart [INRIA, Intern, from Feb 2023 until Aug 2023]
- Grégoire Dhimoila [ENS DE LYON, Intern, from Jun 2023 until Aug 2023]
- Corentin Duvivier [UNIV LILLE, Intern, from Jun 2023 until Aug 2023]
- Abdellah El Mrini [INRIA, Intern, from May 2023 until Sep 2023]
- Xuan Vinh Ho [INRIA, Intern, from Feb 2023 until Jul 2023]
- Ismail Labiad [INRIA, Intern, from Mar 2023 until Aug 2023]
- Basile Tschora [INRIA, Intern, from Aug 2023 until Sep 2023]

Administrative Assistant

- Aurore Dalle [INRIA]

External Collaborator

- Rémi Gilleron [UNIV LILLE, HDR]

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 smart-

phones, 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 new research project FedMalin (see Section 10.4.4) is dedicated to this objective for the Federated Learning setting.

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

6.1 Awards

- Edwige Cyffers received the L'Oréal-Unesco prize for her work on privacy-preserving machine learning.
- Aurélien Bellet was promoted as research director at INRIA.
- Jan Ramon received a best newcomer award for the work [13].

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

Keyword: Natural language processing

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

Contact: Pascal Denis

Participant: Pascal Denis

Partner: Orange Labs

7.1.2 Mangoes

Name: MAgnet liNGuistic wOrd vEctorS

Keywords: Word embeddings, NLP

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: Aurelien 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:

- builds on abstractions general enough to write backbone algorithmic code agnostic to the actual computation framework, statistical model details or network communications setup - 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 - 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...) - 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: - the client, to be run by each participant, performs the learning on local data et releases only the result of the computation - the server orchestrates the process and aggregates the local models in a global model

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

Contact: Aurelien Bellet

Participants: Paul Andrey, Aurelien 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

Authors: Gaurav Maheshwari, Michael Perrot

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

We publicly release all datasets accompanying publications. Moreover, we actively participate in initiatives promoting dataset availability and transparency.

Data Provenance Initiative Damien Sileo co-founded the Data Provenance Initiative (DPI), a project focused on transparency about the natural language processing datasets used for the final stages of large language models training. This stage, known as fine-tuning, involves converting a multitude of existing datasets into input-output pairs for language modeling. The DPI addresses a key oversight in this process: the neglect of metadata during dataset aggregation. One notable achievement of DPI is leveraging [37] to recover missing licenses and attributions, revealing inaccuracies in licenses on prominent platforms like GitHub and HuggingFace (the most popular hub for NLP datasets). DPI also traced the origins of texts and annotated datasets with additional details like date, task types, and geographical origins, facilitating comprehensive audits of datasets in natural language processing.

- Contact person: Damien Sileo
- Datasets: https://www.dataprovenance.org/ (and the associated github link available at this address)

Tasksource Commitment to public dataset release is essential for open and reproducible science. However, datasets often require preprocessing for effective model training or evaluation. The Tasksource project [39] enhances dataset usability by offering one-line commands for loading datasets into standardized formats, ensuring smooth training and evaluation. Tasksource introduces a system for annotating preprocessing steps, provides numerous preprocessing annotations, and has uploaded over a hundred datasets to HuggingFace/tasksource, sourced from GitHub, academic websites, or directly from authors.

- Contact person: Damien Sileo
- Datasets: https://github.com/sileod/tasksource

8 New results

8.1 Natural Language Processing

Exploring Category Structure with Contextual Language Models and Lexical Semantic Networks [29]

Recent work on predicting category structure with distributional models, using either static word embeddings (Heyman and Heyman, 2019) or contextualized language models (CLMs) (Misra et al., 2021), report low correlations with human ratings, thus calling into question their plausibility as models of human semantic memory. In this work, we revisit this question testing a wider array of methods for probing CLMs for predicting typicality scores. Our experiments, using BERT (Devlin et al., 2018), show the importance of using the right type of CLM probes, as our best BERT-based typicality prediction methods substantially improve over previous works. Second, our results highlight the importance of polysemy in this task: our best results are obtained when using a disambiguation mechanism. Finally, additional experiments reveal that Information Contentbased WordNet (Miller, 1995), also endowed with disambiguation, match the performance of the best BERT-based method, and in fact capture complementary information, which can be combined with BERT to achieve enhanced typicality predictions.

A Tale of Two Laws of Semantic Change: Predicting Synonym Changes with Distributional Semantic Models [23]

Lexical Semantic Change is the study of how the meaning of words evolves through time. Another related question is whether and how lexical relations over pairs of words, such as synonymy, change over time. There are currently two competing, apparently opposite hypotheses in the historical linguistic literature regarding how synonymous words evolve: the Law of Differentiation (LD) argues that synonyms tend to take on different meanings over time, whereas the Law of Parallel Change (LPC) claims that synonyms tend to undergo the same semantic change and therefore remain synonyms. So far, there has been little research using distributional models to assess to what extent these laws apply on historical corpora. In this work, we take a first step toward detecting whether LD or LPC operates for given word pairs. After recasting the problem into a more tractable task, we combine two linguistic resources to propose the first complete evaluation framework on this problem and provide empirical evidence in favor of a dominance of LD. We then propose various computational approaches to the problem using Distributional Semantic Models and grounded in recent literature on Lexical Semantic Change detection. Our best approaches achieve a balanced accuracy above 0.6 on our dataset. We discuss challenges still faced by these approaches, such as polysemy or the potential confusion between synonymy and hypernymy.

WordNet Is All You Need: A Surprisingly Effective Unsupervised Method for Graded Lexical Entailment [28]

We propose a simple unsupervised approach which exclusively relies on WordNet (Miller, 1995) for predicting graded lexical entailment (GLE) in English. Inspired by the seminal work of Resnik (1995), our method models GLE as the sum of two information-theoretic scores: a symmetric semantic similarity score and an asymmetric specificity loss score, both exploiting the hierarchical synset structure of WordNet. Our approach also includes a simple disambiguation mechanism to handle polysemy in a given word pair. Despite its simplicity, our method achieves performance above the state of the art (Spearman $\rho = 0.75$) on HyperLex (Vulic et al., 2017), the largest GLE dataset, outperforming all previous methods, including specialized word embeddings approaches that use WordNet as weak supervision.

Find-2-Find: Multitask Learning for Anaphora Resolution and Object Localization [27]

In multimodal understanding tasks, visual and linguistic ambiguities can arise. Visual ambiguity can occur when visual objects require a model to ground a referring expression in a video without strong supervision, while linguistic ambiguity can occur from changes in entities in action flows. As an example from the cooking domain, "oil" mixed with "salt" and "pepper" could later be referred to as a "mixture". Without a clear visual-linguistic alignment, we cannot know which among several objects shown is referred to by the language expression "mixture", and without resolved antecedents, we cannot pinpoint what the mixture is. We define this chicken-and-egg problem as visual-linguistic ambiguity. In this paper, we present Find2Find, a joint anaphora resolution and object localization dataset targeting the problem of visual-linguistic ambiguity, consisting of 500 anaphora-annotated recipes with corresponding videos. We present experimental results of a novel end-to-end joint multitask learning framework for Find2Find that fuses visual and textual information and shows improvements both for anaphora resolution and object localization as compared to a strong single-task baseline.

Probing neural language models for understanding of words of estimative probability [31]

Words of Estimative Probability (WEP) are phrases used to express the plausibility of a statement. Examples include terms like probably, maybe, likely, doubt, unlikely, and impossible. Surveys have shown that human evaluators tend to agree when assigning numerical probability levels to these WEPs. For instance, the term highly likely equates to a median probability of 0.90±0.08 according to a survey by Fagen-Ulmschneider (2015). In this study, our focus is to gauge the competency of neural language processing models in accurately capturing the consensual probability level associated with each WEP. Our first approach is utilizing the UNLI dataset (Chen et al., 2020), which links premises and hypotheses with their perceived joint probability p. From this, we craft prompts in the form: "[PREMISE]. [WEP], [HYPOTHESIS]." This allows us to evaluate whether language models can predict if the consensual probability level of a WEP aligns closely with p. In our second approach, we develop a dataset based on WEP-focused probabilistic reasoning to assess if language models can logically process WEP compositions. For example, given the prompt "[EVENTA] is likely. [EVENTB] is impossible.", a wellfunctioning language model should not conclude that [EVENTA&B] is likely. Through our study, we observe that both tasks present challenges to out-of-the-box English language models. However, we also demonstrate that fine-tuning these models can lead to significant and transferable improvements.

MindGames: Targeting Theory of Mind in Large Language Models with Dynamic Epistemic Modal Logic [40]

Theory of Mind (ToM) is a critical component of intelligence, yet accurately measuring it continues to be a subject of debate. Prior research has attempted to apply human ToM assessments to natural language processing models using either human-created standardized tests or rule-based templates. However, these methods primarily focus on simplistic reasoning and require further validation. In this study, we utilize dynamic epistemic logic, which has established overlaps with ToM, to generate more intricate problems. We also introduce novel verbalization techniques to express these problems using natural language. Our findings indicate that certain language model scaling (from 70M to 6B and 350M to 174B) does not consistently yield results better than random chance. While GPT-4 demonstrates improved epistemic reasoning capabilities, there is still room for enhancement. Our code and datasets are publicly available here and there.

tasksource: A Dataset Harmonization Framework for Streamlined NLP Multi-Task Learning and Evaluation [39]

The HuggingFace Datasets Hub hosts thousands of datasets, offering exciting opportunities for language model training and evaluation. However, datasets for a specific task type often have different schemas, making harmonization challenging. Multi-task training or evaluation necessitates manual work to fit data into task templates. Several initiatives independently tackle this issue by releasing harmonized datasets or providing harmonization codes to preprocess datasets into a consistent format. We identify patterns across previous preprocessing efforts, such as column name mapping and extracting specific

sub-fields from structured data in a column. We then propose a structured annotation framework that ensures our annotations are fully exposed and not hidden within unstructured code. We release a dataset annotation framework and dataset annotations for more than 500 English tasks[1]. These annotations include metadata, such as the names of columns to be used as input or labels for all datasets, which can save time for future dataset preprocessing, regardless of whether our framework is utilized. We fine-tune a multi-task text encoder on all tasksource tasks, outperforming every publicly available text encoder of comparable size in an external evaluation.

Models of Modals [32]

Modal verbs in English communicate delicate shades of meaning, there being a large range of verbs both on the necessity side (must, have to, should, ought to, need, need to) and the possibility side (can, may, could, might, be able to). They therefore constitute excellent test ground to apply and compare different methodologies that can lay bare the factors that drive the speaker's choice of modal verb. This book is not merely concerned with a purely grammatical description of the use of modal verbs, but aims at advancing our understanding of lexical and grammatical units in general and of linguistic methodologies to explore these. It thus involves a genuine effort to compare, assess and combine a variety of approaches. It complements the leading descriptive qualitative work on modal verbs by testing a diverse range of quantitative methods, while not ignoring qualitative issues pertaining to the semantics-pragmatics interface. Starting from a critical assessment of what constitutes the meaning of modal verbs, different types of empirical studies (usage-based, data-driven and experimental), drawing considerably on the same data sets, shows how method triangulation can contribute to an enhanced understanding. Due attention is also given to individual variation as well as the degree to which modals can predict L2 proficiency level.

8.2 Data Sets

The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI [37]

The race to train language models on vast, diverse, and inconsistently documented datasets has raised pressing concerns about the legal and ethical risks for practitioners. To remedy these practices threatening data transparency and understanding, we convene a multidisciplinary effort between legal and machine learning experts to systematically audit and trace 1800+ text datasets. We develop tools and standards to trace the lineage of these datasets, from their source, creators, series of license conditions, properties, and subsequent use. Our landscape analysis highlights the sharp divides in composition and focus of commercially open vs closed datasets, with closed datasets monopolizing important categories: lower resource languages, more creative tasks, richer topic variety, newer and more synthetic training data. This points to a deepening divide in the types of data that are made available under different license conditions, and heightened implications for jurisdictional legal interpretations of copyright and fair use. We also observe frequent miscategorization of licenses on widely used dataset hosting sites, with license omission of 70%+ and error rates of 50%+. This points to a crisis in misattribution and informed use of the most popular datasets driving many recent breakthroughs. As a contribution to ongoing improvements in dataset transparency and responsible use, we release our entire audit, with an interactive UI, the Data Provenance Explorer, which allows practitioners to trace and filter on data provenance for the most popular open source finetuning data collections: www.dataprovenance.org.

8.3 Decentralized Learning

Refined Convergence and Topology Learning for Decentralized SGD with Heterogeneous Data [20]

One of the key challenges in decentralized and federated learning is to design algorithms that efficiently deal with highly heterogeneous data distributions across agents. In this paper, we revisit the analysis of the popular Decentralized Stochastic Gradient Descent algorithm (D-SGD) under data heterogeneity. We exhibit the key role played by a new quantity, called neighborhood heterogeneity, on the convergence rate of D-SGD. By coupling the communication topology and the heterogeneity, our analysis sheds light

[1] See here

on the poorly understood interplay between these two concepts. We then argue that neighborhood heterogeneity provides a natural criterion to learn data-dependent topologies that reduce (and can even eliminate) the otherwise detrimental effect of data heterogeneity on the convergence time of D-SGD. For the important case of classification with label skew, we formulate the problem of learning such a good topology as a tractable optimization problem that we solve with a Frank-Wolfe algorithm. As illustrated over a set of simulated and real-world experiments, our approach provides a principled way to design a sparse topology that balances the convergence speed and the per-iteration communication costs of D-SGD under data heterogeneity.

One-Shot Federated Conformal Prediction [22]

In this paper, we introduce a conformal prediction method to construct prediction sets in a oneshot federated learning setting. More specifically, we define a quantile-of-quantiles estimator and prove that for any distribution, it is possible to output prediction sets with desired coverage in only one round of communication. To mitigate privacy issues, we also describe a locally differentially private version of our estimator. Finally, over a wide range of experiments, we show that our method returns prediction sets with coverage and length very similar to those obtained in a centralized setting. Overall, these results demonstrate that our method is particularly well-suited to perform conformal predictions in a one-shot federated learning setting.

8.4 Learning and Speech Recognition

Differentially private speaker anonymization [18]

Sharing real-world speech utterances is key to the training and deployment of voice-based services. However, it also raises privacy risks as speech contains a wealth of personal data. Speaker anonymization aims to remove speaker information from a speech utterance while leaving its linguistic and prosodic attributes intact. State-of-the-art techniques operate by disentangling the speaker information (represented via a speaker embedding) from these attributes and re-synthesizing speech based on the speaker embedding of another speaker. Prior research in the privacy community has shown that anonymization often provides brittle privacy protection, even less so any provable guarantee. In this work, we show that disentanglement is indeed not perfect: linguistic and prosodic attributes still contain speaker information. We remove speaker information from these attributes by introducing differentially private feature extractors based on an autoencoder and an automatic speech recognizer, respectively, trained using noise layers. We plug these extractors in the state-of-the-art anonymization pipeline and generate, for the first time, differentially private utterances with a provable upper bound on the speaker information they contain. We evaluate empirically the privacy and utility resulting from our differentially private speaker anonymization approach on the LibriSpeech data set. Experimental results show that the generated utterances retain very high utility for automatic speech recognition training and inference, while being much better protected against strong adversaries who leverage the full knowledge of the anonymization process to try to infer the speaker identity.

8.5 Privacy

From Noisy Fixed-Point Iterations to Private ADMM for Centralized and Federated Learning [21]

We study differentially private (DP) machine learning algorithms as instances of noisy fixed-point iterations, in order to derive privacy and utility results from this well-studied framework. We show that this new perspective recovers popular private gradient-based methods like DP-SGD and provides a principled way to design and analyze new private optimization algorithms in a flexible manner. Focusing on the widely-used Alternating Directions Method of Multipliers (ADMM) method, we use our general framework to derive novel private ADMM algorithms for centralized, federated and fully decentralized learning. For these three algorithms, we establish strong privacy guarantees leveraging privacy amplification by iteration and by subsampling. Finally, we provide utility guarantees using a unified analysis that exploits a recent linear convergence result for noisy fixed-point iterations.

DP-SGD Without Clipping: The Lipschitz Neural Network Way [36]

State-of-the-art approaches for training Differentially Private (DP) Deep Neural Networks (DNN) faces difficulties to estimate tight bounds on the sensitivity of the network's layers, and instead rely on a process of per-sample gradient clipping. This clipping process not only biases the direction of gradients but also proves costly both in memory consumption and in computation. To provide sensitivity bounds and bypass the drawbacks of the clipping process, our theoretical analysis of Lipschitz constrained networks reveals an unexplored link between the Lipschitz constant with respect to their input and the one with respect to their parameters. By bounding the Lipschitz constant of each layer with respect to its parameters we guarantee DP training of these networks. This analysis not only allows the computation of the aforementioned sensitivities at scale but also provides leads on to how maximize the gradient-to-noise ratio for fixed privacy guarantees. To facilitate the application of Lipschitz networks and foster robust and certifiable learning under privacy guarantees, we provide a Python package that implements building blocks allowing the construction and private training of such networks.

GAP: Differentially Private Graph Neural Networks with Aggregation Perturbation [30]

In this paper, we study the problem of learning Graph Neural Networks (GNNs) with Differential Privacy (DP). We propose a novel differentially private GNN based on Aggregation Perturbation (GAP), which adds stochastic noise to the GNN's aggregation function to statistically obfuscate the presence of a single edge (edge-level privacy) or a single node and all its adjacent edges (node-level privacy). Tailored to the specifics of private learning, GAP's new architecture is composed of three separate modules: (i) the encoder module, where we learn private node embeddings without relying on the edge information; (ii) the aggregation module, where we compute noisy aggregated node embeddings based on the graph structure; and (iii) the classification module, where we train a neural network on the private aggregations for node classification without further querying the graph edges. GAP's major advantage over previous approaches is that it can benefit from multi-hop neighborhood aggregations, and guarantees both edge-level and node-level DP not only for training, but also at inference with no additional costs beyond the training's privacy budget. We analyze GAP's formal privacy guarantees using Rényi DP and conduct empirical experiments over three real-world graph datasets. We demonstrate that GAP offers significantly better accuracy-privacy trade-offs than state-of-the-art DP-GNN approaches and naive MLP-based baselines. Our code is publicly available here.

High-Dimensional Private Empirical Risk Minimization by Greedy Coordinate Descent [25]

In this paper, we study differentially private empirical risk minimization (DP-ERM). It has been shown that the worst-case utility of DP-ERM reduces polynomially as the dimension increases. This is a major obstacle to privately learning large machine learning models. In high dimension, it is common for some model's parameters to carry more information than others. To exploit this, we propose a differentially private greedy coordinate descent (DP-GCD) algorithm. At each iteration, DP-GCD privately performs a coordinate-wise gradient step along the gradients' (approximately) greatest entry. We show theoretically that DP-GCD can achieve a logarithmic dependence on the dimension for a wide range of problems by naturally exploiting their structural properties (such as quasi-sparse solutions). We illustrate this behavior numerically, both on synthetic and real datasets.

Differential Privacy has Bounded Impact on Fairness in Classification [26]

We theoretically study the impact of differential privacy on fairness in classification. We prove that, given a class of models, popular group fairness measures are pointwise Lipschitz-continuous with respect to the parameters of the model. This result is a consequence of a more general statement on accuracy conditioned on an arbitrary event (such as membership to a sensitive group), which may be of independent interest. We use this Lipschitz property to prove a non-asymptotic bound showing that, as the number of samples increases, the fairness level of private models gets closer to the one of their non-private counterparts. This bound also highlights the importance of the confidence margin of a model on the disparate impact of differential privacy.

Rényi Pufferfish Privacy: General Additive Noise Mechanisms and Privacy Amplification by Iteration via Shift Reduction Lemmas [38]

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 machine learning 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. We also derive stronger guarantees against out-ofdistribution adversaries. 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.

Exploiting Problem Structure in Privacy-Preserving Optimization and Machine Learning [33]

In the past decades, concerns about the societal impact of machine learning have been growing. Indeed, if machine learning has proven its usefulness in science, day-to-day applications, and many other domains, its success is principally due to the availability of large datasets. This raises two concerns, the first about the confidentiality of the training data, and the second, about possible discrimination in a model's predictions. Trustworthy machine learning aims at providing technical answers to these concerns.Unfortunately, guaranteeing the privacy of the training data and the fairness of the predictions often decreases the utility of the learned model. This problem has drawn significant interest in the past years, but most of existing methods (usually based on stochastic gradient descent) tend to fail in some common scenarios, like training of high-dimensional models. In this thesis, we study how structural properties of machine learning problems can be exploited to improve the trade-off between privacy and utility, and how this can impact the fairness of the predictions.The first two contributions of this thesis are two new differentially private optimization algorithms, that are both based on coordinate descent. They aim at exploiting different structural properties of the problem at hand. The first algorithm is based on stochastic coordinate descent, and can exploit imbalance in the scale of the gradient's coordinates by using large step sizes. This allows our algorithm to obtain useful models in difficult problems, where stochastic gradient descent quickly stalls. The second algorithm is based on greedy coordinate descent. Its greedy updates allow to focus on the most important coordinates of the problem, which can sometimes drastically improve utility (e.g. when the solution of the problem is sparse).The third contribution of this thesis studies the interplay of differential privacy and fairness in machine learning. These two notions have rarely been studied simultaneously, and there are growing concerns that differential privacy may exacerbate unfairness. We show that group fairness measures have interesting regularity properties, provided that the predictions of the model are Lipschitz-continuous in its parameters. This result allows to derive a bound on the difference in fairness levels between a private model and its non-private counterpart.

Privacy-preserving Learning by Averaging in Collaborative Networks [34]

In recent years, due to the growing importance of network applications and the growing concerns for privacy, there is an increasing interest in decentralized forms of machine learning. In this dissertation, we study the setting that involves a communication network of agents, where each agent locally privatizes (adds noise to) its data, and where the agents aim to collaboratively learn statistical models over their data. Such local privatization is in line with a standard of data privacy known as local differential privacy, and local differential privacy is useful when alternatives, such as secure multi-party computation or central differential privacy performed by a trusted curator, are infeasible. However, local differential privacy results, typically, in worse utility (less accurate statistical models) compared to central differential privacy because, for the same privacy budget, local differential privacy adds more privatization noise than central differential privacy. The principal question of this dissertation is the following: given that local differential privacy must be used, how could the agents maximize the utility they achieve? We study two cases to address the stated principal question.In the first case, we consider the problem of

distributed averaging, where each agent intends to collaboratively compute the unbiased average over the individual values of all agents without revealing neither their sensitive attributes nor their degree (number of neighbors). Usually, existing works solve this problem by assuming that either (i) each agent reveals its degree to its neighbors or (ii) every two neighboring agents can perform handshakes (requests that rely on replies) in every exchange of information. Since such assumptions are not always desirable, we propose an approach that is handshake-free and where the degrees are privatized. In particular, we use a gossip algorithm that computes averages that are biased when the graph of agents is non-regular (when the vertices have unequal degrees) and then perform a procedure combining multiple biased averages for bias correction. We apply the proposed approach for fitting linear regression models. We prove the asymptotic guarantee that the mean squared error between the average of privatized attributes computed by our approach and the average of sensitive attributes is $\mathcal{O}\left(\frac{1}{n}\right)$, where n is the number of agents.In the second case, we consider a group of agents, where features (for fitting regression models) are computed by transforming sensitive attributes, and where the transformations have high-magnitude gradients or singularities. In such setting, there is a risk to magnify the privatization noise if the perturbed data is in an interval where the feature function has high-magnitude gradients. We provide a tailored noise mechanism for privatizing features by solving a convex program in such a way that (i) only pertinent intervals of transformations are selected, (ii) the variance of privatization noise is minimized, and (iii) the biasedness of privatization noise is minimized.

8.6 Fairness and Transparency

Introducing the TRUMPET project: TRUstworthy Multi-site Privacy Enhancing Technologies [19]

This paper is an overview of the EU-funded project TRUMPET (see here), and gives an outline of its scope and main technical aspects and objectives. In recent years, Federated Learning has emerged as a revolutionary privacy-enhancing technology. However, further research has cast a shadow of doubt on its strength for privacy protection. 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, crossdomain, cross-border European datasets with privacy guarantees that follow 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.

Fair Without Leveling Down: A New Intersectional Fairness Definition [24]

In this work, we consider the problem of intersectional group fairness in the classification setting, where the objective is to learn discrimination-free models in the presence of several intersecting sensitive groups. First, we illustrate various shortcomings of existing fairness measures commonly used to capture intersectional fairness. Then, we propose a new definition called the α-Intersectional Fairness, which combines the absolute and the relative performance across sensitive groups and can be seen as a generalization of the notion of differential fairness. We highlight several desirable properties of the proposed definition and analyze its relation to other fairness measures. Finally, we benchmark multiple popular in-processing fair machine learning approaches using our new fairness definition and show that they do not achieve any improvement over a simple baseline. Our results reveal that the increase in fairness measured by previous definitions hides a "leveling down" effect, i.e., degrading the best performance over groups rather than improving the worst one.

FairGrad: Fairness Aware Gradient Descent [14]

We tackle the problem of group fairness in classification, where the objective is to learn models that do not unjustly discriminate against subgroups of the population. Most existing approaches are limited to simple binary tasks or involve difficult to implement training mechanisms. This reduces their practical applicability. In this paper, we propose FairGrad, a method to enforce fairness based on a reweighting scheme that iteratively learns group specific weights based on whether they are advantaged or not. FairGrad is easy to implement and can accommodate various standard fairness definitions. Furthermore,

we show that it is comparable to standard baselines over various datasets including ones used in natural language processing and computer vision.

Private Sampling with Identifiable Cheaters [16]

In this paper we study verifiable sampling from probability distributions in the context of multi-party computation. This has various applications in randomized algorithms performed collaboratively by parties not trusting each other. One example is differentially private machine learning where noise should be drawn, typically from a Laplace or Gaussian distribution, and it is desirable that no party can bias this process. In particular, we propose algorithms to draw random numbers from uniform, Laplace, Gaussian and arbitrary probability distributions, and to verify honest execution of the protocols through zero-knowledge proofs. We propose protocols that result in one party knowing the drawn number and protocols that deliver the drawn random number as a shared secret.

8.7 Machine Learning

A Revenue Function for Comparison-Based Hierarchical Clustering [15]

Comparison-based learning addresses the problem of learning when, instead of explicit features or pairwise similarities, one only has access to comparisons of the form: *Object A is more similar to B than to C*. Recently, it has been shown that, in Hierarchical Clustering, single and complete linkage can be directly implemented using only such comparisons while several algorithms have been proposed to emulate the behaviour of average linkage. Hence, finding hierarchies (or dendrograms) using only comparisons is a well understood problem. However, evaluating their meaningfulness when no ground-truth nor explicit similarities are available remains an open question. In this paper, we bridge this gap by proposing a new revenue function that allows one to measure the goodness of dendrograms using only comparisons. We show that this function is closely related to Dasgupta's cost for hierarchical clustering that uses pairwise similarities. On the theoretical side, we use the proposed revenue function to resolve the open problem of whether one can approximately recover a latent hierarchy using few triplet comparisons. On the practical side, we present principled algorithms for comparison-based hierarchical clustering based on the maximisation of the revenue and we empirically compare them with existing methods.

8.8 Applications in the Health Domain

Automated detection of toxicophores and prediction of mutagenicity using PMCSFG algorithm [17]

Maximum common substructures (MCS) have received a lot of attention in the chemoinformatics community. They are typically used as a similarity measure between molecules, showing high predictive performance when used in classification tasks, while being easily explainable substructures. In the present work, we applied the Pairwise Maximum Common Subgraph Feature Generation (PMCSFG) algorithm to automatically detect toxicophores (structural alerts) and to compute fingerprints based on MCS. We present a comparison between our MCS-based fingerprints and 12 well-known chemical fingerprints when used as features in machine learning models. We provide an experimental evaluation and discuss the usefulness of the different methods on mutagenicity data. The features generated by the MCS method have a state-of-the-art performance when predicting mutagenicity, while they are more interpretable than the traditional chemical fingerprints.

A Smartphone-based Architecture for Prolonged Monitoring of Gait [35]

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.

8.9 Theoretical Computer Science

Linear Programs with Conjunctive Database Queries [13]

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.

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 International initiatives

10.1.1 Participation in other International Programs

SLANT: Bilateral ANR project with Luxembourg

Participants: Pascal Denis *(contact person)*, Aurélien Bellet, Mikaela Keller, Gaurav Maheshwari.

Acronym: SLANT

Title: Spin and bias in Language Analyzed in News and Texts

Duration: December 2019 – December 2023

Coordinator: Philippe Muller, IRIT, Toulouse

Partners: IRIT (Toulouse), SnT (Luxembourg)

Abstract: There is a growing concern about misinformation or biased information in public communication, whether in traditional media or social forums. While automating fact-checking has received a lot of attention, the problem of fair information is much larger and includes more insidious forms like biased presentation of events and discussion. The SLANT project aims at characterizing bias in textual data, either intended, in public reporting, or unintended in writing aiming at neutrality. An abstract model of biased interpretation using work on discourse structure, semantics and interpretation will be complemented and concretized by finding relevant lexical, syntactic, stylistic or rhetorical differences through an automated but explainable comparison of texts with different biases on the same subject, based on a dataset of news media coverage from a diverse set of sources. We will also explore how our results can help alter bias in texts or remove it from automated representations of texts.

IMPRESS: Bilateral Inria-DFKI project

Participants: Pascal Denis *(contact person)*, Rémi Gilleron, Priyansh Trivedi.

Acronym: IMPRESS

Title: Improving Embeddings with Semantic Knowledge

Duration: October 2020-February 2024

Coordinator: PASCAL DENIS and IVANA KRUIJFF-KORBAYOVA (DFKI)

Partners: Sémagramme (Inria Nancy), DFKI (Germany)

Abstract: Virtually all NLP systems nowadays use vector representations of words, a.k.a. word embeddings. Similarly, the processing of language combined with vision or other sensory modalities employs multimodal embeddings. While embeddings do embody some form of semantic relatedness, the exact nature of the latter remains unclear. This loss of precise semantic information can

affect downstream tasks. Furthermore, while there is a growing body of NLP research on languages other than English, most research on multimodal embeddings is still done on English. The goals of IMPRESS are to investigate the integration of semantic knowledge into embeddings and its impact on selected downstream tasks, to extend this approach to multimodal and mildly multilingual settings, and to develop open source software and lexical resources, focusing on video activity recognition as a practical testbed.

10.2 International research visitors

10.2.1 Visits to international teams

Research stays abroad

Tudor Cebere

Visited institution: Vector Institute

Country: Canada

Dates: Sept. 23 - Aug. 24

Context of the visit: Collaboration with the team of Nicolas Papernot

Mobility program/type of mobility: internship

Viet-Toan Le

Visited institution: AMAAI (Audio, Music and AI Lab) of SUTD (Singapore University of Technology and Design)

Country: Singapore

Dates: October 14th to 24th

Context of the visit: Programme Hubert Curien

Mobility program/type of mobility: research stay

10.3 European initiatives

10.3.1 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
- TIME.LEX (time.lex), 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
- UNIVERSIDAD DE VIGO (UVIGO), Spain
- ARTEEVO TECHNOLOGIES LTD (ARTEEVO), Israel

Inria contact: Jan Ramon

Coordinator: Ruth Muleiro Alonso. Fundación Centro Tecnolóxico de Telecomunicacións de Galicia

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
- TIME.LEX (time.lex), Belgium
- TECHNOVATIVE SOLUTIONS LTD, United Kingdom
- HL7 INTERNATIONAL FONDATION (HL7 INTERNATIONAL), Belgium
- FUNDACION CENTRO TECNOLOXICO DE TELECOMUNICACIONS DE GALICIA (GRADIANT), Spain
- 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

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.

10.4 National initiatives

10.4.1 ANR DEEP-Privacy (2019–2023)

Participants: Marc Tommasi *(contact person)*, Aurélien Bellet, Pascal Denis, Jan Ramon, Brij Mohan Lal Srivastava, Rishabh Gupta.

DEEP-PRIVACY proposes a new paradigm based on a distributed, personalized, and privacy-preserving approach for speech processing, with a focus on machine learning algorithms for speech recognition. To this end, we propose to rely on a hybrid approach: the device of each user does not share its raw speech data and runs some private computations locally, while some cross-user computations are done by communicating through a server (or a peer-to-peer network). To satisfy privacy requirements at the acoustic level, the information communicated to the server should not expose sensitive speaker information.

10.4.2 HyAIAI. INRIA Defi (2019-2023)

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

HyAIAI is an Inria Defi about the design of novel, interpretable approaches for Artificial Intelligence.

Recent progress in Machine Learning (ML) and especially Deep Learning has made ML pervasive in a wide range of applications. However, current approaches rely on complex numerical models: their decisions, as accurate as they may be, cannot be easily explained to the layman that may depend on

these decisions (ex: get a loan or not). In the HyAIAI IPL, we tackle the problem of making "Interpretable ML" through the study and design of hybrid approaches that combine state of the art numeric models with explainable symbolic models. More precisely, our goal is to be able to integrate high level (domain) constraints in ML models, to give model designers information on ill-performing parts of the model, and to give the layman/practitioner understandable explanations on the results of the ML model.

10.4.3 ANR PMR (2020-2024)

Participants: Jan Ramon *(contact person)*, Aurélien Bellet, Marc Tommasi, Cesar Sabater.

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

10.4.4 FedMalin. INRIA Defi (2021-2024)

Participants: Aurélien Bellet *(contact person)*, Jan Ramon, Marc Tommasi, Michaël Perrot, Batiste Le Bars, Edwige Cyffers, Paul Mangold, Tudor Cebere.

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.

10.4.5 FLAMED: Federated Learning and Analytics on Medical Data. INRIA Action Exploratoire (2020-2024)

Participants: Aurélien Bellet *(contact person)*, Marc Tommasi, Paul Mangold, Nathan Bigaud, Paul André.

Flamed is about decentralized approaches for AI in medicine. The main objective is to operate data analysis and machine learning tasks in a network of hospital units without any data exchange. This approach helps to solve data privacy and sovereignty issues while taking advantage of the statistical power of federation and collaboration. This research is done in collaboration with the Lille Hospital.

10.4.6 ANR-JCJC PRIDE (2020–2025)

Participants: Aurélien Bellet *(contact person)*, Marc Tommasi, Jan Ramon, Edwige Cyffers, Batiste Le Bars, Paul Mangold, Tudor Cebere.

Machine learning (ML) is ubiquitous in AI-based services and data-oriented scientific fields but raises serious privacy concerns when training on personal data. The starting point of PRIDE is that personal data should belong to the individual who produces it. This requires to revisit ML algorithms to learn from many decentralized personal datasets while preventing the reconstruction of raw data. Differential Privacy (DP) provides a strong notion of protection, but current decentralized ML algorithms are not able to learn useful models under DP. The goal of PRIDE is to develop theoretical and algorithmic tools that enable differentially-private ML methods operating on decentralized datasets, through two complementary objectives: (1) prove that gossip protocols naturally reinforce DP guarantees; (2) propose algorithms at the intersection of decentralized ML and secure multi-party computation.

10.4.7 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).

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

Participants: Aurélien Bellet *(contact person)*, Jan Ramon, Marc Tommasi, Michaël Perrot, Cesar Sabater, Edwige Cyffers, Paul Mangold, Tudor Cebere.

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.

10.4.9 CAPS'UL (2023-2028)

Participant: Marc Tommasi *(contact person)*.

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.

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

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

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.

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

Participants: Jan Ramon *(contact person)*, Marc Tommasi, Michaël Perrot, Arnaud Descours, Vitalii Emilianov.

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.

10.5 Regional initiatives

10.5.1 STARS: Fairness in decentralized and privacy-preserving machine learning (2021-2023).

Participant: Michaël Perrot *(contact person)*.

The aim of this project is to propose approaches capable of learning a non-discriminatory model in a context where data is distributed among several entities that wish to preserve the confidentiality of the data in their possession.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

Member of the organizing committees

- Aurélien Bellet co-organizes the Federated Learning One World webinar (1100+ registered attendees) since May 2020

11.1.2 Scientific events: selection

- Marc Tommasi served as Area Chair for ECML-PKDD and ICLR and PC member for UAI (top reviewer).
- Aurélien Bellet served as Area Chair for ICML and NeurIPS.
- Michaël Perrot served as Reviewer for NeurIPS (Top Reviewer) and ICML
- Jan Ramon was PC member of AFT@NeurIps'23, AISTATS'24, BNAIC'23, DGM4G@NeurIps'23, DS'23, ECML/PKDD'23, ICBINB@NeurIps'23, ICLR'24, ICML'23, IJCAI'23, MLG@ECML/PKDD'23, MLG@KDD'23, NeurIps'23, PPAI@AAAI'23, SDM'24, XKDD@ECML/PKDD'23.

- MIKAELA KELLER served as Reviewer for *SEM'23 and CAp'23.
- DAMIEN SILEO served as Reviewer for ACL Rolling Reviews.
- PASCAL DENIS was Action Editor for ACL Rolling Review, and served as PC member for ACL'23, EACL'23, and CODI@ACL'23.

11.1.3 Journal

- AURÉLIEN BELLET is Action Editor for Transactions of Machine Learning Research (TMLR).
- JAN RAMON is member of the editorial boards of Machine Learning Journal (MLJ), Data Mining and Knowledge Discovery (DMKD), Journal of Machine Learning Research (JMLR), ECML-PKDD Journal track. JAN RAMON is action editor of Data Mining and Knowledge Discovery (DMKD).
- PASCAL DENIS is standing reviewer for Transactions of the Association for Computational Linguistics (TACL).

11.1.4 Invited talks

- MICHAËL PERROT gave a talk titled "Differential Privacy has Bounded Impact on Fairness in Classification" at the MICS Seminars.
- DAMIEN SILEO gave a talk named "Generative AI and Large Language Models" at the Cyle Supérieur du Numérique.
- AURÉLIEN BELLET gave a talk at the Privacy and Fairness in AI for Health workshop, the Paris Privacy-Preserving ML workshop and the Privacy Alpine Seminar (Privaski).
- MARC TOMMASI gave a talk at the days on Distributed Learning organized by the GDR RSD in association with SIF.

11.1.5 Scientific expertise

- MARC TOMMASI was a member (scientific expert) of the recruitment committee of full professors at Lens and Marseille.
- AURÉLIEN BELLET was a member of the recruitement committee of an assistant professor at Université Jean Monnet de Saint-Etienne.
- MIKAELA KELLER was a member of the recruitement committees of an assistant professor at Université Jean Monnet de Saint-Etienne and of an assistant professor at Université de Lille.
- AURÉLIEN BELLET serves as ethics advisor for the ESFRI project SLICES-PP.
- JAN RAMON was reviewer for ANR, COST, Horizon Europe and ChistERA.

11.1.6 Research administration

- JAN RAMON was a member of CER (Commission Emploi Recherche) in the INRIA Center of Lille University.
- AURÉLIEN BELLET is member of the Operational Committee for the assessment of Legal and Ethical risks (COERLE).
- MARC TOMMASI is co-head of the DatInG group (4 teams, about 100 persons), member of the Conseil Scientifique du laboratoire CRIStAL and member of the Commission mixte CRIStAL/Faculty of Science, Lille University.
- 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).

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- Licence MIASHS: MARC TOMMASI, Data Science, 24h, L2, Université de Lille.
- Licence Informatique: MARC TOMMASI, Introduction to AI, 24h, L2, Université de Lille.
- Licence MIASHS: MIKAELA KELLER, Data Science, 12h, L2, Université de Lille.
- Licence MIASHS: MIKAELA KELLER, Data Science 2, 24h, L3, Université de Lille.
- Licence MIASHS: MIKAELA KELLER, Traitement de données, 24h, L2, Université de Lille.
- Master MIASHS: MIKAELA KELLER, Algorithmes fondamentaux de la fouille de données, 27h, M1, Université de Lille.
- Master Data Science: MIKAELA KELLER, Machine Learning 1, 24h, M1, 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 MIASHS: MICHAËL PERROT, Algorithmes fondamentaux de la fouille de données, 27h, M1, 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 Data Science: MARC TOMMASI Seminars 24h.
- Master Data Science: AURÉLIEN BELLET, Privacy Preserving Machine Learning, 24h, M2, Université de Lille and Ecole Centrale de Lille.
- Parcours Science des Données et Intelligence Artificielle: AURÉLIEN BELLET, Advanced Machine Learning, 6h, Ecole Centrale de Lille.
- Training in Privacy-Preserving and Federated Machine Learning, 3 days, researchers from L'Oréal R&D.
- Master Sciences Cognitives: DAMIEN SILEO, Machine Learning for Cognitive Sciences, 8h, M2, Université de Lille.
- 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. Since Nov. 2022. MICHAËL PERROT.
- Postdoc: BATISTE LE BARS. On collaboration graph design for decentralized learning. October 2021-Jun. 2023. AURÉLIEN BELLET and MARC TOMMASI
- Postdoc: ARNAUD DESCOURS. On federated optimization with lower communication cost. Since Nov. 2023. JAN RAMON
- Postdoc: BAPTISTE COTTIER. On sparse secure aggregation. May-Sept 2023. JAN RAMON

- Postdoc: SABRI CHELLOUF. On privacy-preserving learning for mobile devices. May 2023 - Dec 2023. JAN RAMON
- Postdoc: IMANE TALBI. On the relation between statistical privacy and security assumptions. May 2023-April 2024. JAN RAMON
- PhD defended in Feb. 2023: NICOLAS CROSETTI, Privacy Risks of Aggregates in Data Centric-Workflows. FLORENT CAPELLI and SOPHIE TISON and JOACHIM NIEHREN and JAN RAMON.
- Phd defended in Jun. 2023: ARIJUS PLESKA, Tractable Probabilistic Models for Large Scale Networks [34]. JAN RAMON
- PhD in progress: MOITREE BASU, Integrated privacy-preserving AI, since 2019. JAN RAMON.
- PhD defended in Oct. 2023: PAUL MANGOLD. Decentralized Optimization and privacy [33]. AURÉLIEN BELLET and MARC TOMMASI and JOSEPH SALMON, since October 2020.
- Phd in progress: GAURAV MAHESHWARI. Trustworthy Representations for Natural Language Processing, since Nov 2020. AURÉLIEN BELLET, MIKAELA KELLER, and PASCAL DENIS
- Phd in progress: EDWIGE CYFFERS. Decentralized learning and privacy amplification, since Oct. 2021. AURÉLIEN BELLET
- 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: TUDOR CEBERE. Privacy-Preserving Machine Learning, since Nov. 2022. AURÉLIEN BELLET
- 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 Privacy and Fairness, since Sept. 2023, AURÉLIEN BELLET, CATUSCIA PALAMIDESSI and MICHAËL PERROT.
- Engineer SOPHIE VILLEROT, ADT project Tailed: Trustworthy AI Library for Environments which are Decentralized, since Nov. 2020. JAN RAMON
- Engineer KAMALKUMAR MACWAN, ADT project Tailed: Trustworthy AI Library for Environments which are Decentralized, Jan-Oct. 2023. JAN RAMON

- Engineer JOSEPH RENNER, Improving Word Representations with Semantic Knowledge, Nov. 2020-Oct 2023. PASCAL DENIS and RÉMI GILLERON
- Engineer RISHABH GUPTA, Disentanglement approaches for speech data. Apr. 2023-Jul. 2023. AURÉLIEN BELLET and MARC TOMMASI.
- Engineer PAUL ANDREY, Decentralized and Federated Learning with DecLearn. Since Jul. 2022. AURÉLIEN BELLET and MARC TOMMASI.
- Engineer NATHAN BIGAUD, Decentralized and Federated Learning with DecLearn. Oct. 2022-Oct.2023.AURÉLIEN BELLET and MARC TOMMASI.
- Engineer QUENTIN SINH. Integrating MAGNET results in the TRUMPET privacy-preserving federated learning platform, Since Nov 2023. JAN RAMON.
- Engineer KEVIN NGAKOSSA. Integrating MAGNET results on constraint-based privacy in the TRUMPET privacy-preserving federated learning platform, Since Oct. 2023. JAN RAMON.
- Engineer LÉONARD DEROOSE. Development of TRUMPET privacy-preserving platform components, Since Sep. 2023. JAN RAMON.
- Engineer LI MOU. Statistical privacy computation algorithms, Since Feb. 2023. JAN RAMON.

11.2.3 Juries

- MARC TOMMASI member of the PhD jury of Victor Connes (Reviewer), Othmane Marfoq (Reviewer), Eduardo Brandao (President).
- MICHAËL PERROT: member of the PhD jury of Ganesh del Grosso (Examiner)
- JAN RAMON: member of the PhD jury of Angelo Saadeh, Les Applications du Calcul Multipartite Sécurisé en Apprentissage Automatique (Examiner)
- JAN RAMON: member of the PhD jury of Arnaud Griver-Sibert, Combining differential privacy and homomorphic encryption for privacy-preserving collaborative machine learning (Reviewer)
- JAN RAMON member of the PhD jury of Juan Alvarado, Modeling relational structures via constraint graphons (dept of computer science KULeuven). (Co-director)
- AURÉLIEN BELLET: member of the PhD jury of Marina Constantini (Examiner)

11.3 Popularization

11.3.1 Articles and contents

- MICHAËL PERROT: Article on the Inria Website on “AI decentralized: how to ensure more fairness and privacy?”

11.3.2 Interventions

- MICHAËL PERROT: Pint of Science “IA d’la joie” on fairness in artificial intelligence, Médiathèque de la Madeleine, France.
- MICHAËL PERROT: Utopiades organized by the Alter/Echos association on fairness in artificial intelligence, Remotely.

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. '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.

[4] 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.

[5] 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.

[6] 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.

[7] C. Pelekis, J. Ramon and Y. Wang. 'Hölder-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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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

[13] F. Capelli, N. Crosetti, J. Niehren and J. Ramon. 'Linear Programs with Conjunctive Database Queries'. In: *Logical Methods in Computer Science* (3rd Jan. 2024). URL: https://hal.science/hal-04317553.

[14] G. Maheshwari and M. Perrot. 'FairGrad: Fairness Aware Gradient Descent'. In: *Transactions on Machine Learning Research Journal* (Aug. 2023). URL: https://hal.science/hal-03902196.

[15] A. Mandal, M. Perrot and D. Ghoshdastidar. 'A Revenue Function for Comparison-Based Hierarchical Clustering'. In: *Transactions on Machine Learning Research Journal* (Apr. 2023). URL: https://hal.science/hal-03902209.

[16] C. Sabater, F. Hahn, A. Peter and J. Ramon. 'Private Sampling with Identifiable Cheaters'. In: *Proceedings on Privacy Enhancing Technologies* 2023.2 (2023). URL: https://inria.hal.science/hal-03904200.

[17] L. Schietgat, B. Cuissart, K. D. Grave, K. Efthymiadis, R. Bureau, B. Crémilleux, J. Ramon and A. Lepailleur. 'Automated detection of toxicophores and prediction of mutagenicity using PMCSFG algorithm'. In: *Molecular Informatics* (2023). DOI: 10.1002/minf.202200232. URL: https://hal.science/hal-03940446.

[18] 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). DOI: 10.48550/arXiv.2202.11823. URL: https://inria.hal.science/hal-03588932.

Invited conferences

[19] A. Pedrouzo-Ulloa, J. Ramon, P. Duflot, F. Pérez-González, S. Lilova, Z. Chihani, N. Gentili, P. Ulivi, M. A. Hoque, T. Mukammel, Z. Pritzker, A. Lemesle, J. Loureiro-Acuña, X. Martínez and G. Jiménez-Balsa. 'Introducing the TRUMPET project: TRUstworthy Multi-site Privacy Enhancing Technologies'. In: IEEE CSR 2P-DPA workshop - Workshop on Privacy-Preserving Data Processing and Analysis. Venice, Italy, 31st July 2023. URL: https://inria.hal.science/hal-04092216.

International peer-reviewed conferences

[20] B. L. Bars, A. Bellet, M. Tommasi, E. Lavoie and A.-M. Kermarrec. 'Refined Convergence and Topology Learning for Decentralized SGD with Heterogeneous Data'. In: Proceedings of The 26th International Conference on Artificial Intelligence and Statistics (AISTATS 2023). Valencia, Spain, Spain, 2023. URL: https://inria.hal.science/hal-03905091.

[21] E. Cyffers, A. Bellet and D. Basu. 'From Noisy Fixed-Point Iterations to Private ADMM for Centralized and Federated Learning'. In: Proceedings of the 40th International Conference on Machine Learning (ICML). Honolulu, United States, July 2023. URL: https://hal.science/hal-04260417.

[22] P. Humbert, B. Le Bars, A. Bellet and S. Arlot. 'One-Shot Federated Conformal Prediction'. In: ICML 2023 - 40th International Conference on Machine Learning. Proceedings of the 40th International Conference on Machine Learning (ICML). Honolulu (Hawai), United States, 23rd July 2023. URL: https://hal.science/hal-03981605.

[23] B. Liétard, M. Keller and P. Denis. 'A Tale of Two Laws of Semantic Change: Predicting Synonym Changes with Distributional Semantic Models'. In: *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics*. The 12th Joint Conference on Lexical and Computational Semantics. Toronto, Canada, June 2023. URL: https://inria.hal.science/hal-04126662.

[24] G. Maheshwari, A. Bellet, P. Denis and M. Keller. 'Fair Without Leveling Down: A New Intersectional Fairness Definition'. In: EMNLP 2023 - The 2023 Conference on Empirical Methods in Natural Language Processing. Singapore (SG), Singapore, 6th Dec. 2023. URL: https://hal.science/hal-04273353.

[25] P. Mangold, A. Bellet, J. Salmon and M. Tommasi. 'High-Dimensional Private Empirical Risk Minimization by Greedy Coordinate Descent'. In: *International Conference on Artificial Intelligence and Statistics*. AISTATS 2023 - International Conference on Artificial Intelligence and Statistics. Valencia, Spain, 25th Apr. 2023. URL: https://inria.hal.science/hal-03714465.

[26] P. Mangold, M. Perrot, A. Bellet and M. Tommasi. 'Differential Privacy has Bounded Impact on Fairness in Classification'. In: *Proceedings of the 40th International Conference on Machine Learning*. International Conference on Machine Learning. Vol. 202. Honolulu, United States, 23rd July 2023. URL: https://hal.science/hal-03902203.

[27] C. Oguz, P. Denis, E. Vincent, S. Ostermann and J. van Genabith. 'Find-2-Find: Multitask Learning for Anaphora Resolution and Object Localization'. In: 2023 Conference on Empirical Methods in Natural Language Processing. Singapore, Singapore, 2023. URL: https://hal.science/hal-04259861.

[28] J. Renner, P. Denis and R. Gilleron. 'WordNet Is All You Need: A Surprisingly Effective Unsupervised Method for Graded Lexical Entailment'. In: Findings of the Association for Computational Linguistics: EMNLP 2023. Singapore, France, 2023. URL: https://hal.science/hal-04250849.

[29] J. Renner, P. Denis, R. Gilleron and A. Brunellière. 'Exploring Category Structure with Contextual Language Models and Lexical Semantic Networks'. In: *Exploring Category Structure with Contextual Language Models and Lexical Semantic Networks*. EACL 2023 - 17th Conference of the European Chapter of the Association for Computational Linguistics. Dubrovnik, Croatia, 13th May 2023. URL: https://inria.hal.science/hal-03986142.

[30] S. Sajadmanesh, A. S. Shamsabadi, A. Bellet and D. Gatica-Perez. 'GAP: Differentially Private Graph Neural Networks with Aggregation Perturbation'. In: USENIX Security 2023 - 32nd USENIX Security Symposium. Anaheim, United States, 9th Aug. 2023. URL: https://inria.hal.science/hal-03905068.

[31] D. Sileo and M.-F. Moens. 'Probing neural language models for understanding of words of estimative probability'. In: Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023). Toronto, France: Association for Computational Linguistics, July 2023, pp. 469–476. DOI: 10.18653/v1/2023.starsem-1.41. URL: https://hal.science/hal-04290243.

Scientific books

[32] I. Depraetere, B. Cappelle, M. Hilpert, L. D. Cuypere, M. Dehouck, P. Denis, S. Flach, N. Grabar, C. Grandin, T. Hamon, C. Hufeld, B. Leclercq and H.-J. Schmid. *Models of Modals: From Pragmatics and Corpus Linguistics to Machine Learning*. Vol. Topics in English Linguistics [TiEL]. 110. De Gruyter Mouton, 2023. URL: https://hal.science/hal-03984358.

Doctoral dissertations and habilitation theses

[33] P. Mangold. 'Exploiting Problem Structure in Privacy-Preserving Optimization and Machine Learning'. Université de Lille, 11th Oct. 2023. URL: https://hal.science/tel-04346432.

[34] A. Pleska. 'Privacy-preserving Learning by Averaging in Collaborative Networks'. Université de Lille, 6th June 2023. URL: https://hal.science/tel-04198205.

Reports & preprints

[35] L. Bart, E. A. Bechorfa, A. Boutet, J. Ramon and C. Frindel. *A Smartphone-based Architecture for Prolonged Monitoring of Gait*. Insa Lyon; Inria Lyon, 20th Dec. 2023. URL: https://hal.science/hal-04355370.

[36] L. Béthune, T. Masséna, T. Boissin, C. Friedrich, F. Mamalet, A. Bellet, M. Serrurier and D. Vigouroux. *DP-SGD Without Clipping: The Lipschitz Neural Network Way*. 25th May 2023. URL: https://hal.science/hal-04130913.

[37] S. Longpre, R. Mahari, A. Chen, N. Obeng-Marnu, D. Sileo, W. Brannon, N. Muennighoff, N. Khazam, J. Kabbara, K. Perisetla, A. Wu, E. Shippole, K. Bollacker, T. Wu, L. Villa, S. Pentland and S. Hooker. *The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI*. 4th Nov. 2023. URL: https://hal.science/hal-04290233.

[38] 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*. 21st Dec. 2023. URL: https://inria.hal.science/hal-04363020.

[39] D. Sileo. *tasksource: A Dataset Harmonization Framework for Streamlined NLP Multi-Task Learning and Evaluation*. 17th May 2023. URL: https://inria.hal.science/hal-04099649.

[40] D. Sileo and A. Lernould. *MindGames: Targeting Theory of Mind in Large Language Models with Dynamic Epistemic Modal Logic*. 5th May 2023. URL: https://inria.hal.science/hal-04098588.