Proposal for a Ph.D. topic

# Collaborative Multimodal Learning

**Ph.D. Supervisors:** Younès Bennani, Prof. Institut Galilée, LIPN-CNRS, Younes.Bennani@sorbonne-paris-nord.fr
Basarab Matei, MCF-HDR, LIPN-CNRS, Basarab.Matei@sorbonne-paris-nord.fr

**Hosting research team:** ADA@A3, LIPN UMR 7030 CNRS

## Research context

The proposed work is in the continuity of one of the major axes developed by the ADA component of the A3 team of LIPN for many years on unsupervised learning, and also in the continuity of research activities recently developed related to deep learning, self-supervised learning, and learning of representations.

## Topic Description

Data is frequently presented in a format that spans many modalities, such as video material, which normally comprises at least an image and sound (voice, music), as well as text in the form of subtitles, which are often in a language other than the spoken data. Other situations, such as human sensing, provide data in several modalities, such as face expression data in the form of photographs mixed with auditory (voice, sound), haptic (touch), or other sensory data. The Multimodal Data Analysis research challenge is thus focused on the integration and interpretation of data within and across modalities, as well as human engagement with multimodal data and the knowledge and insights gained from it. The outcomes from this research will enable improved understanding and modeling of rich data sources in domains such as business, health, environment, and education. Learning how to describe and summarize multimodal data in a way that makes use of the complementarity and redundancy of various modalities is the fundamental problem. The variety of multimodal data makes creating such representations difficult. Language, for example, is frequently symbolic, but audio and visual modalities are represented as signals. Due to the heterogeneity of the data, various issues arise naturally, such as different types of noise, modalities (or viewpoints) alignment, and missing data approaches. We look at multimodal representations from joint perspectives: projecting all modalities to a single space while maintaining information from the given modalities. The features must fulfill several conditions at least be descriptive, summarized, and stable, for instance, extracting information from high-dimensional and noisy datasets. We propose to use a field based on a very robust mathematical theory, that guarantees the relevance of the information that we will extract and prove many interesting qualities of this information, especially stability and multi-scaling. This approach is built on the concept that the shape of datasets contains pertinent information, which leads us to use invariants which is a tool that can capture topological changes across the entire range of scales and store this information into a coherent clear and easy to represent data. Generally, the hypothesis behind is that the characteristics that persist for a wide range of parameters are "true" characteristics for data.

## Objective

In this thesis, we propose to conduct a study to explore the discovery and definition of common spaces of multimodal representations. In this space of common topological representations, we propose to create coherence between the different data modalities through a collaborative learning process. This common representation will be proposed to deep learning systems in order to enrich the inputs and improve the performance of these systems. Our goal now is to apply a rigorous theory to improve and ameliorate machine learning models, this can be achieved by vectorising the topological features from multimodal data or by transforming the model for example by designing specialized layers of neural network that are capable of handling such features. So, we will either, introduce to models the information extracted using topology and compare the result and the stability with other outcomes using other types of information extracting, or we will combine topological information with other classical descriptors to achieve more. We aim to answer the following questions:
**Q1** How represent each example as a topological features for all types of multimodal data?
**Q2** Is the stability of the method will mix up classes if they are so closes?
**Q3** Is the distance used in comparing the topological features like bottleneck distance that compares barcodes efficient and discriminating or it's better to compare after vectorization?
**Q4** Which is better vectorising the topological features from multimodal data or adjusting the neural network model? or it depends? and depends on what?
**Q5** How to create coherence between the different data modalities through a collaborative learning process?

## Methodology

First, investigate the state-of-the-art of existing solutions on multimodal learning. Second, propose a first approach to the problem validate and compare it to the state-of-the-art methods. Then apply it to real-life application.

## Expected contributions and outreach

We expect from this PhD proposal different types of contributions:

- **Scientific contributions in terms of journal papers and conference communications** in the field of Machine learning and data science, corresponding both to new learning-based schemes and Topological Data Analysis;
- **Algorithms and codes** using deep learning frameworks (eg, tensorflow, pytorch), which we expect to be of interest to develop an open source collaborative multimodal learning toolbox;
- **Datasets and associated benchmarking experiments** which will be used to distribute a data challenge.

## Desired skills

The targeted PhD candidate shall have a MSc and/or engineer degree in Data Science and applied mathematics with a strong interest in Machine Learning, possibly acknowledged by previous activities or experience. A dual degree in Machine Learning and data science as promoted by EID$^2$ MSc program would be of key interest. Besides a strong theoretical background, computer skills, including first experience in using state-of-the-art deep learning frameworks (e.g., tensorflow, pytorch) and programming environment (e.g., python, git server), will be particularly expected.