

## OFFRE DE STAGE DE MASTER (français et anglais)

Titre : Approche Machine Learning pour les études de nanocomposites

### ENCADREMENT DU STAGE

**Encadrant principal** : Delong HE et Filippo Gatti

**Co-encadrants**: Jinbo BAI, Didier Clouteau

**Contact** : [delong.he@centralesupelec.fr](mailto:delong.he@centralesupelec.fr), [filippo.gatti@centralesupelec.fr](mailto:filippo.gatti@centralesupelec.fr)

**Lieu du stage** : Laboratoire MSSMAT, CentraleSupélec, 3 Rue Joliot Curie, 91190, Gif-sur-Yvette

**Profil(s) de candidats souhaité(s)** : Master 2 Mathématiques Appliquées et/ou Matériaux numériques ou Elève ingénieur– Intérêt pour l'intelligence artificielle.

---

### Contexte :

L'utilisation des technologies avancées en microscopie électronique et optique a permis d'éclairer les relations entre les propriétés des matériaux et leurs microstructures multi-échelles (allant des molécules aux échelles micro ou millimétriques). Cependant, actuellement, nous employons encore principalement la méthode essai-erreur, lorsqu'on optimise une composition de matériau ou un processus d'ingénierie, en suivant le paradigme composition-processus-structure-propriété. Les données ainsi obtenues et nos connaissances sont fragmentées, malgré des caractérisations structurales à différentes échelles. Cet inconvénient empêche une compréhension globale de la corrélation complexe entre les structures des matériaux à différentes échelles, en particulier dans le cas des nanocomposites (polymères avec nanocharges).

La méthode *data-driven* qui a émergé récemment, fournit un ensemble d'outils robustes pour l'étude de phénomènes multi-échelles et multi-physiques. Les méthodes d'apprentissage automatique (ML) sont de plus en plus appliquées en science des matériaux pour concevoir de nouveaux matériaux ou pour optimiser certaines de leurs propriétés spécifiques. Le ML tire parti de grands ensembles de données étiquetés, telles des images haute résolution, pour en extraire des informations significatives. Ces données représentent les caractéristiques de l'ensemble des données en jeu, ou, en d'autres termes, des régularités cachées. Un algorithme de ML correctement adapté - formé sur de grands ensembles de données - peut donc apprendre à mapper des données brutes (par exemple des images de microscopie) en données mesurables (module de Young, conductivité électrique, etc.).

### Objectifs :

Ce stage a pour but d'intégrer l'approche ML dans les études de composites polymères comportant des nanocharges. Les principaux travaux consistent à : (1) construire une base de données des microstructures de composites; et (2) concevoir un algorithme de ML efficace et le tester sur la base de données préalablement assemblée.

Une base de données fiable (comme CIFAR10 classique, CIFAR100, MNIST, etc.) peut inclure différentes images de microscopie. Certaines données déjà publiées sur ces matériaux seront également incluses. La conception d'un algorithme ML sera essentielle pour extraire des caractéristiques cachées significatives et générales, pour estimer la distribution des données et pour prédire les propriétés macroscopiques des matériaux. Compte tenu du grand nombre d'outils de ML

prêts à l'emploi disponibles de nos jours, une grande partie de l'effort sera consacrée à la définition d'une architecture appropriée, ainsi qu'au choix des connaissances d'entrée dont l'algorithme de ML peut bénéficier.

---

## English Version :

### Title: Machine learning used for nanocomposites research

#### Context:

The application of advanced microscopy technology (electron and optical) sheds light on the relationships between material properties and their multi-scale microstructures (ranging from molecules to micro or millimeter scales). However, nowadays, a trial & error approach is still dominated in seeking a material composition and corresponding engineering process by following the paradigm *composition-process-structure-property*. One of the bottlenecks of this research approach is represented by an overall poor and scattered knowledge, despite insightful multi-scale structural characterizations. This drawback prevents the global understanding of the complex correlation among material structures at different scales, particularly in the case of nanocomposites (polymers with nanofillers).

data-driven modelling methods are emerging as a robust tool-set for the study of multi-scale, multi-physics phenomena. Machine Learning (ML) methods are increasingly being applied in materials science to either design new materials or to either optimize certain specific properties. ML takes advantage of large labelled data sets - such as high-resolution images - to extract meaningful *features* out of them. Those *features* represent modal characteristics of the data set at stake, or, in other words, hidden regularities that can be interpreted as or that can relate to specific physical or chemical phenomena. A properly tailored ML algorithm - trained on large data sets - can therefore learn to map raw and meaningless data (e.g. microscopy images) into insightful and measurable features (Young's modulus, electrical conductivity etc).

#### Description of the work:

The objective of this study is to develop a ML-aided material performance enhancement framework, targeting at polymer composites with nanofillers. The main works consist of: (1) constructing a database of local material microstructures; and (2) designing an effective ML algorithm to be trained upon the experimental database previously assembled.

A reliable database for multi-purpose ML research (such as classical CIFAR10, CIFAR100, MNIST etc) may primarily include different microscopy images. Some of already-published data on such materials will be included. The conception of a ML algorithm will be essential to extract meaningful and general hidden *features*, to estimate data distribution and to predict – once trained – macroscopic interrelated material properties. Given the large number of *ready-made* ML tools available nowadays, much of the effort will be consecrated to the definition of an appropriate architecture, as well as to the choice of input knowledge the ML algorithm can benefit from.

**The expected outcome** is twofold: the compilation of a state-of-art material database and the design of a powerful ML algorithm trained upon it.