# Utilizing machine learning to enhance efficiency of large-scale data based interatomic potential

Molecules' combination exploration is essential in many industrial fields by driving the innovation process. One particular accurate example would be the pharmaceutical industry which is always in a race for discovering new drugs. Even though Physics' laws dictate the possible atomic combinations regarding the interatomic potentials, the universe of viable molecules remains huge. Exploring it all seems as hard and efficiently as Sisyphus' work. It clearly appears that more clever ways of exploration have to be followed. Such ways can be obtained through machine learning.

Machine learning represents a group of artificial algorithms that are trained to find patterns in the input datasets. Therefore, the data has to represent accurately the subject to seek accuracy and effectiveness. It implies to own the a good amount of data properly spread within the field of work.

While several correlations exist to determine interatomic potentials, the first of them used to be determined manually. The time and work required to do so is huge. In the meantime, a spectacular improvement of computing performances and of machine learning occurred. By dividing the method to determine interatomic potentials in three steps, machine learning algorithms can be used. They are listed as follows :

1. Build a dataset
2. Model the atomic environment
3. Use regression methods

Such methods, if well used, are time saving. However, the dependence on the initial dataset implies that the deduced correlations cannot be used outside of the dataset's range without errors.

Allegro is a tool developed for the pharmaceutical industry. It is based of a large dataset, accessible through MPNN. It has shown encouraging results for the discovering of drug molecules. In order to run it, 8 expensive GPU are used : theses algorithms are still reserved for important industrial companies.