Title of the thesis	Artificial Intelligence for High-Throughput Catalysts Design
Acronym	AICAT
Reference number	019

Hosting institution	**Employer**
Centrale Lille Website: https://centralelille.fr/	Centrale Lille Website: https://centralelille.fr/
Hosting research unit 1	**Hosting research unit 2**
Name: Unit of Catalysis and Solid State Chemistry Acronym: UCCS Identification number: UMR 8181 Address: Université de Lille Campus Cité scientifique Bâtiment C3 59655 Villeneuve d'Ascq Cedex – France Website: http://uccs.univ-lille1.fr/index.php/fr/	Name: Centre de Recherche en Informatique, Signal et Automatique de Lille Acronym: CRIStAL Identification number: UMR 9189 Address: Université de Lille Campus scientifique Bâtiment ESPRIT Avenue Henri Poincaré 59655 Villeneuve d'Ascq Website: https://www.cristal.univ-lille.fr/
Principal supervisor	**Co-supervisor**
Name: Sébastien Surname: PAUL Email: sebastien.paul@centralelille.fr Phone: +33 (0)3 20 33 54 57	Name: Pierre Surname: CHAINAIS Email: pierre.chainais@centralelille.fr Phone: +33 (0)3 28 77 84 51

Thesis information

Keywords	Artificial Intelligence, Machine Learning, High-Throughput Technologies, Hybrid Catalysis, Biorefineries.
Abstract	In chemistry, the discovery of new reactions and processes often relies on a big amount of exploratory experiments inspired by a lot of expert knowledge. The best strategy is often a compromise between *exploitation* of existing expert knowledge and *exploration* of new settings and formulas. This experimental design is even more crucial when searching for new catalysts. Catalysts are material that tend to favour and accelerate certain chemical reactions, sometimes by a tremendous factor. Applications are omnipresent, with a strong impact on crucial domains such as environment, food, health, energy, to cite a few. The objective of the AICAT project is to develop a machine learning approach capable of establishing rapidly the best choice of the input parameters of the problem, e.g. the formula and synthesis conditions or the physicochemical properties of the catalysts, to reach the main objective, e.g. the highest activity or yield into the molecules of interest. The design of a new catalyst comprises the systematic exploration of sample libraries that must be synthesized, analysed and tested for a targeted reaction of interest. High-throughput catalytic screening uses robots to take advantage of the parallelization and automation of these operations. This is precisely what REALCAT (Advanced High-Throughput Technologies Platform for Biorefineries Catalysts Design, www.realcat.fr) permits to carry out. REALCAT, located at Centrale Lille, is managed by the University of Lille and coordinated by the Unit of Catalysis and Solid State Chemistry (UCCS UMR 8181,

	http://uccs.univ-lille1.fr) with a set of partners including CRIStAL laboratory (UMR 9189, https://www.cristal.univ-lille.fr) and the Charles Viollette Institute (ICV UMR INRA 1281, http://institutcharlesviollette.com). This platform is able to synthesize catalysts as well as to test their performance under reaction conditions and to characterize them at high-speed on a large set of robots. Any kind of catalyst (chemo- or bio-catalysts and even the novel concept of hybrid catalysts combining both types) can be considered on REALCAT. The development chain is gathered on the same site, which makes this platform unique in the world. The automated exploration generates a large amount of data that must then be formatted, analysed and exploited to optimize the quest for new catalysts and protocols. The objective of the AICAT project is to develop a machine learning approach to the large data set that the system can provide. A better understanding of the input-output relation of the platform will be exploited to speed up the discovery of the optimal catalyst for a given reaction through the rational selection of the most efficient synthesis conditions, taking the best benefit from REALCAT power. The project aims at exploiting the richness of existing data from previous and current experiments to optimize the future behaviour of REALCAT through an efficient experimental design. The PhD student will be co-supervised by experienced researchers from UCCS and CRIStAL labs. During the course of the PhD a secondment will be done in Shanghai, China, in the E2P2 laboratory, which is a mixed unit between CNRS and the SOLVAY company.
Expected profile of the candidate	We are looking for a graduate student with a Master level in statistics or computer science or signal processing; a Master level in chemistry with a good knowledge of machine learning and programming could also be suitable. In this multidisciplinary context, the PhD student must have strong skills in data processing and machine learning, but also an interest in green chemistry and catalysis. A good culture in a subset of the following fields is expected: statistics, optimization, reinforcement learning, experimental design, programming, data bases….
Application procedure	The application procedure is detailed on the European programme PEARL website www.pearl-phd-lille.eu. The funding is managed by the I-SITE ULNE foundation which is a partnership foundation between the University of Lille, Engineering schools like Centrale Lille, research organisms, the Institut Pasteur de Lille and the University hospital. The application file will have to be submitted before April 15, 2020 (10h Paris Time) and emailed to the following address : international@isite-ulne.fr.
Net salary and Lump Sum	A net salary of about €1,600 + €530 per month to cover mobility, travel and family costs.