3iA Côte d'Azur
Interdisciplinary Institute for Artificial Intelligence

Ph.D. research topic

- Title of the proposed topic: Identification of molecular patterns to inform about chemical safety
- Research axis of the 3IA: Core elements of AI + AI for computational biology and bio-inspired IA
- **Supervisor (name, affiliation, email):** We propose a co-supervision between a chair and an affiliate chair : Pr. Vincent Vandewalle (UCA/Inria, Chair) Vincent.vandewalle@inria.fr and Dr. David Rouquié (Bayer, Affiliate Chair) david.rouquie@bayer.com
- Potential co-supervisor (name, affiliation):

The laboratory and/or research group: MAASAI Inria joint team-project with LJAD & I3S / Bayer Crop science toxicology department

---

Apply by sending an email directly to the supervisor.
The application will include:

- Letter of recommendation of the supervisor indicated above
- Curriculum vitæ.
- Motivation Letter.
- Academic transcripts of a master's degree(s) or equivalent.
- At least, one letter of recommendation.
- Internship report, if possible.

⇨ **All the requested documents must be gathered and concatenated in a single PDF file named in the following format: LAST NAME of the candidate_Last Name of the supervisor_2023.pdf**

- Description of the topic:

  The proposed research PhD project will be based on the use of a large public data set already generated in the project JUMP-CP (https://jump-cellpainting.broadinstitute.org/) and a new dataset that will be generated in an international consortium called OASIS (Omics for Assessing Signatures for Integrated Safety) a join HESI/Broad Institute Led Consortium that is just being initiated (early 2023) . The Health and Environmental Sciences Institute (HESI) a non-profit institution whose mission is to collaboratively identify and help to resolve global health and environmental challenges through the engagement of scientists from academia, government, industry, NGOs, and other strategic partners. Bayer is a

member of HESI and is actively contributing to different projects and has initiated the OASIS project. Dr. David Rouquié is the Program Co-chair together with Dr. Anne Carpenter at the Broad Institute.

Overall, in OASIS we propose a new collaborative initiative to advance a novel data-driven approach utilizing image-based profiling for toxicological applications. The objective of the consortium is to gain confidence in the combination of Cell Painting and transcriptomics for safety assessment using hepatotoxicity as use-case. This effort supports the drive to shift from chemical/drug safety decisions based on apical animal endpoints to in vitro testing and predictive toxicology methods that may support the reduction in the volume and/or duration of animal testing while enhancing mechanistic insights.

The agrochemical and pharmaceutical industries share a mission to deliver safe and effective chemicals to society. Animal testing provides some assurance, but it is expensive, slow, and often fails to accurately predict impact on humans, as evidenced in the pharma industry by clinical trial failure rates due to unexpected toxicity: 30% phase 1 and 25% phase 2 (Fabre et al 2019; Harrison et al. 2016). Regulatory agencies worldwide are now pushing to Replace, Reduce, and Refine or even ban animal testing; new policies have sparked a paradigm shift towards alternative testing methods that are more accurate for human safety assessment while less reliant on animal testing. Efficient and reliable novel approaches that replace or reduce animal testing, while increasing the safety and efficacy of new agrochemical and medicines products, are urgently needed to meet this demand. In OASIS project 10 pharmaceutical/agrochemical and 4 technology companies are bring together to create a novel dataset to evaluate several cell models and several high-dimensional readouts to better predict liver toxicity in rats and, ultimately, humans.

We have identified Cell Painting and transcriptomics as promising candidates for this effort because they generate high density biological information while being broadly available and cost-effective. While these technologies have been shown to reflect compound-induced modulation of diverse targets and biochemical processes, the useful characteristics for an integrative evaluation of chemical bioactivities must be coupled with a point of departure (POD) coupled with exposure estimated to be used in chemical safety assessment. These technologies have been shown to capture compound-induced modulation of diverse targets and biochemical processes, and to report on mechanisms of toxicity; however, they are not well-validated, particularly for liver toxicity assessment. The project is targeting two distinct but converging applications: support of the generation of novel insights by pooling existing in vivo toxicity data for a large number of data rich compounds (i.e., those already well characterized in animal studies) and comparing these data with novel Cell Painting and transcriptomics data in the context of the anticipated drastic reduction of laboratory animal for safety assessment of agrochemical products while improving preclinical to clinical translatability for pharmaceutical products.

The PhD candidate will use a first-of-its-kind dataset starting with a collection of precious historical/existing data (standardized rat in vivo toxicity studies) from ~200 compounds donated by our pharmaceutical and agrochemical partners, together with ~1500 publicly available compounds and data. We will collect Cell Painting, transcriptomic, and proteomic data for subsets of these compounds in a variety of cell systems that model the liver, ranging from simple cultured cells to primary cells to more complex liver organoids and liver-on-a-chip. Although we do not anticipate creating a perfectly accurate predictor of all types of liver toxicities, we aim to identify which models and readouts can predict which types well, allowing decisions to be made earlier, more accurately, and more cheaply in the development of useful compounds, while reducing animal use.

The objective of this thesis is to develop models capable of inferring the in vivo effect of compounds from in vitro data obtained by Cell Painting. The small number of compounds characterized in vivo makes it difficult to implement supervised learning algorithms to solve this problem. First, to cope with the small number of labeled data we propose to develop unsupervised learning models that can distinguish a finite number of in vitro observed effects induced by the compounds. These models will integrate the dose information which is essential to determine molecular Point of Departure (POD). The groups thus obtained will then be associated with effects observed in vivo in order to establish a correspondence between in vivo and in vitro effects. In a second step, these models will be extended to a semi-supervised framework integrating the available information on in vivo effects and allowing the discovery of new effects in in vitro data. Finally, the whole developed methodology will be extended to take into account all the available information, namely data observed from different cell types and on different species (rats and human).

The successful candidate will thus contribute to:

- Identification of cell painting and gene expression signatures associated with specific hepatoxic profiles observed in the rat and in human
- Perform quantitative analyses to benchmark in vitro molecular Point of Departure (POD) with in vivo No Observed Effect levels determined in vivo