

SPECIES Scholarship Host Application

Topic

The field of Machine Learning (ML) has seen substantial advancements, with an emphasis on interpretability becoming increasingly important (Rudin, 2019). The focus of this project is to explore the use of deep learning to guide symbolic regression, a form of interpretable ML that has been shown to perform on par with or even surpass non-interpretable approaches such as deep learning (La Cava et al., 2021). However, symbolic regression presents unique challenges, including scalability issues with extensive data, where deep learning excels. These hurdles limit the effective application of symbolic regression in environments rich in data.

The aim of this internship is to address these challenges by utilizing deep learning to guide symbolic regression, thereby facilitating its exploration of data not well-represented by existing models. Building on the concept of lexicase selection (Helmuth, 2016), and using the Curiosity-ES algorithm proposed by the Host's team (le Togluenec, 2022), the project will implement deep learning as an exploration bonus for evolution. The ultimate goal is to guide symbolic regression towards sections of data currently underrepresented, thereby facilitating more comprehensive data modeling. This approach presents an exciting avenue for enhancing the interpretability of ML while harnessing the power of deep learning to manage extensive datasets, and holds significant potential for advancing the field.

Helmuth, Thomas, Nicholas Freitag McPhee, and Lee Spector. "Lexicase Selection for Program Synthesis: A Diversity Analysis." *Genetic Programming Theory and Practice XIII* (2016): 151-167.

La Cava, William, et al. "Contemporary Symbolic Regression Methods and their Relative Performance." *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*. 2021.

Le Togluenec, Paul-Antoine, et al. "Curiosity Creates Diversity in Policy Search." *arXiv preprint arXiv:2212.03530* (2022).

Rudin, Cynthia. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* 1.5 (2019): 206-215.

Research Group and Host Institution

This internship will take place at ISAE-SUPAERO in Toulouse, France, in the Learning, Decision, and Optimization (ADO) team in the Department of Complex Systems Engineering. ISAE-SUPAERO is one of the top engineering schools in France and a worldwide leader in aerospace engineering. It is a part of the University of Toulouse and

there will be a strong link in this internship with other research groups in Toulouse, notably the REVA Artificial Life team at IRIT. ISAE-SUPAERO is also member of the Artificial and Natural Intelligence Toulouse Institute. ANITI has been selected as one of four institutes spearheading research on AI in France and aims to make Toulouse one of the world leaders in Artificial Intelligence in research, education, innovation and economic development. ISAE-SUPAERO has recently advanced its research and educational focuses in Artificial Intelligence with the growth of the Decision Systems research group and Science of Decision and Data Master's program.

Over the course of this internship, the student will be able to profit from the lively environment surrounding AI in Toulouse, attending SuReLI, ANITI, and REVA events. The candidate will be surrounded by an international and talented group of students and researchers working on evolutionary computation and machine learning.

The student will also be encouraged to explore Toulouse and its surroundings. Once the capital of the Visigoths and now the capital of the air and space industry, Toulouse offers an excellent way to experience France. It is home to and surrounded by UNESCO World Heritage Sites such as the Basilica of St. Sernin, Carcassonne, and Albi. It is frequently ranked as one of the best student cities and best cities for foreigners. For this internship, we can offer housing on the campus of ISAE-SUPAERO, which is along the beautiful Canal du Midi and easily accessible by bike, metro, or bus.

Advisor

Dennis G. Wilson is an Associate Professor of AI and Data Science at ISAE-SUPAERO in Toulouse, France. They obtained their PhD at the Institut de Recherche en Informatique de Toulouse (IRIT) on the evolution of design principles for artificial neural networks. Prior to that, they worked in the Anyscale Learning For All group in CSAIL, MIT, applying evolutionary strategies and developmental models to the problem of wind farm layout optimization.

Their research focuses on genetic programming, neural networks, and the evolution of learning. They have published in multiple A rank conferences and have co-authored two book chapters on developmental neural networks. They previously organized a competition on wind farm layout optimization at GECCO. They are passionate about environmental applications and are an active member of the Climate Change AI group.