

# Explainability of the decisions taken by an autonomous system

Charles Lesire (ONERA), Arthur Bit-Monnot (LAAS-CNRS)

Autonomy of a system relies greatly on its capacity to take decisions, i.e., to choose which actions to do and how in order to fulfill an objective, reach a goal or carry out a given task. This decision procedure is extremely complex, due to the size of the search space and of non-trivial algorithmic processes (constraint propagation, reification, heuristics, ...). This can make it hard for an end-user to understand why the system chose a particular course of actions to reach its goal.

The objective of this internship is to propose a methodology for explaining the plans produced by a task planning algorithm. The work will take place in the context of hierarchical planning (HTN – hierarchical task networks) and plan-space planning (POP – partial order planning) which provide a number of elements that can be exploited to help explaining the decisions made.

By leveraging prior work carried out at LAAS and ONERA, the intern will analyse the resolution procedure of these algorithms, propose and implement a methodology to explain the final decisions and possibly propose algorithmic changes to improve explainability of decisions. To make the work concrete, the intern will be able to use existing search and rescue scenarios.