# Explicabilité des décisions prises par un système autonome

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

2019-2020

L'autonomie d'un système repose en grande partie sur sa capacité à prendre des décisions, c'est-à-dire à choisir quelles actions réaliser et comment, afin de remplir un objectif, d'atteindre un but, ou de réaliser un certain nombre de tâches. Ce processus de décision est extrêmement complexe, de part la taille de l'espace de recherche, et de plus repose sur des mécanismes complexes (propagation de contraintes, réification, heuristiques, ...) qui peuvent rendre difficile la compréhension par un utilisateur de "pourquoi" le système a choisi ce plan pour réaliser l'objectif donné.

L'objectif de ce stage est de proposer une méthodologie d'explication des plans produits par un algorithme de planification de tâches, en se basant notamment sur les cadres de la planification hiérarchique (par exemple, HTN (Erol, Hendler, and Nau 1994)) et de la planification dans l'espace des plans (ou POP – Partial-Order Planning (Penberthy and Weld 1992)). Ces cadres de planification utilisent des éléments dans la recherche de solutions qui peuvent être utiles à l'explication des décisions prises (Seegebarth et al. 2012).

En s'appuyant sur des travaux en planification hybride menés au LAAS (Dvorak et al. 2014) et à l'ONERA (Bechon et al. 2018), le candidat devra analyser le processus de résolution de ces algorithmes, proposer et implémenter une méthodologie pour expliquer les décisions prises, au regard de la spécification du problème faite par l'utilisateur, et éventuellement proposer des modifications algorithmiques afin d'augmenter l'explicabilité des décisions. Pour concrétiser ce travail, le candidat pourra s'appuyer sur des scénarios de type recherche et sauvetage par des systèmes robotiques autonomes.

Merci de contacter Charles Lesire (`charles.lesire@onera.fr`) et Arthur Bit-Monnot (`bit-monnot@laas.fr`) pour toute information complémentaire. Toute candidature doit être composée à minima d'un CV et d'une lettre de motivation.

# Références


Bechon, Patrick, Magali Barbier, Christophe Grand, Simon Lacroix, Charles Lesire, and Cédric Pralet. 2018. "Integrating planning and execution for a team of heterogeneous robots with time and communication constraints." In *IEEE International Conference on Robotics and Automation (Icra 2018)*. Brisbane, Australia. doi:10.1109/ICRA.2018.8461024.

Dvorak, Filip, Roman Barták, Arthur Bit-Monnot, Félix Ingrand, and Malik Ghallab. 2014. "Planning and Acting with Temporal and Hierarchical Decomposition Models." In *International Conference on Tools with Artificial Intelligence (Ictai 2014)*. Limassol, Cyprus. doi:10.1109/ICTAI.2014.27.

Erol, K., J. Hendler, and D. Nau. 1994. "HTN Planning: Complexity and Expressivity." In *Internatonal Conference on Artificial Intelligence (Aaai 1994)*. Seattle, WA, USA.

Penberthy, Scott, and Daniel Weld. 1992. "UCPOP: A sound, complete, partial order planner for ADL." In *International Conference on Principles of Knowledge Representation and Reasoning (Kr 1992)*. Cambridge, Massachusetts, USA. http://ai.unibo.it/system/files/u11/ucpop-kr92.pdf.

Seegebarth, Bastian, Felix Müller, Bernd Schattenberg, and Susanne Biundo. 2012. "Making Hybrid Plans More Clear to Human Users — a Formal Approach for Generating Sound Explanations." In *International Conference on Automated Planning and Scheduling (Icaps 2012)*. Atibaia, Sao Paulo, Brazil. https://pdfs.semanticscholar.org/552d/c4b225983c166e4a5c6e1250f399a9797fc2.pdf.