Foundations of Trustworthy AI – Integrating Reasoning, Learning and Optimization
TAILOR
Grant Agreement Number 952215

Foundations, techniques, algorithms and tools for allowing autonomous AI agents to decide and learn how to act - D5.1 Report

Document type (nature)	Report
Deliverable No	D5.1
Work package number(s)	5
Date	Due M24, August 31, 2022
Responsible Beneficiary	#6, UNIROMA
Author(s)	Giuseppe De Giacomo
Publicity level	Public
Short description	Report the novel insights, techniques, algorithms and tools developed within the scientific challenges tasks T5.1, T5.2, T5.3, T5.4

History			
Revision	**Date**	**Modification**	**Authors**
v1	2011-11-21	First version	Giuseppe De Giacomo

Document Review		
Reviewer	**Partner ID / Acronym**	**Date of report approval**
Michela Milano	#10/UNIBO	2022-11-08
Philipp Slusallek	#26/DFKI	2022-12-14

Table of Contents

Summary of the report

The main goal of WP5 is to promote the study of foundations, techniques, algorithms and tools for allowing autonomous AI agents to decide and learn how to act. The major challenge is integrating data-based methods with model-based methods by learning first-order symbolic models from non-symbolic data, to allow flexible and compositional reasoning and planning in terms of objects and relations. The interest in particular is to learn meaningful dynamic models from data that allow reasoning and explanation. Apart from the specific scientific work within the project itself, one of the most important objectives of this work-package is to pave the way for research on the topic of "How an AI agent decides and learns how to act" that is multidisciplinary, involving, planning, knowledge representation, synthesis and verification in formal methods, reinforcement learning in non-Markovian models, neuro-symbolic relational methods, and deep learning.

The WP is divided into four scientific challenge tasks, i.e., addressing four main scientific challenges in the theme, plus two extra tasks, one on cross-fertilization with industry and one on fostering a scientific community dedicated to this theme. Each scientific challenge task continuously interacts with the latter two in order to provide input and receive feedback and challenges in order to get a closed-loop approach to the research activities. The scientific challenges will have strong synergies with other WPs.

Introduction to the Deliverable

WP5 has two scientific deliverables D5.1 and D5.2. The first is an intermediate report (M24) and second one is the final report (M48, due at the end of the project). This document is D5.1.

Deliverables

D5.1: Foundations, techniques, algorithms and tools for allowing autonomous AI agents to decide and learn how to act v.1 (report, M24)

D5.2: Foundations, techniques, algorithms and tools for allowing autonomous AI agents to decide and learn how to act v.2 (report, M48)

Organisation

The following people have been involved in the Deliverable:

Partner Acronym	Name	Role
UNIROMA	Giuseppe De Giacomo	Leader of Task 5.1
Universitat Pompeu Fabra	Hector Geffner	Leader of Task 5.2
University of Basel	Malte Helmert	Leader of Task 5.3
Bruno Kessler Institute (FBK)	Paolo Traverso	Leader of Task 5.4
CNRS-IRIT	Andreas Herzig	Leader of Task 5.5
RWTH	Gerhard Lakemeyer	Leader of Task 5.6

Scientific Tasks

The work in WP5 explores the fundamental question: how does an AI agent decide and learn how to act? In particular, WP5 aims at empowering the agent to deliberate autonomously (i.e., without human intervention) how to act in the world. This objective strongly relates to the work in WP4, which focuses on learning how the world works and understanding its properties better.

WP5 aims at realising self-deliberating and autonomous systems by leveraging competencies in Planning and Knowledge Representation and Reasoning, as well as deep competencies in Learning and Optimization. Specifically, WP5 investigates issues like reasoning and planning for acting; learning strategies/plans from data; learning models from data and then do reasoning and planning; learning from past experiences and simulations for refining strategies/plans or models; monitoring the actual outcome of actions; recognizing possibly unexpected outcomes; reasoning, planning and learning how to deal with unexpected outcomes. In general, WP5 explores novel models of world dynamics and agent tasks, a new generation of solvers, and how to integrate data-based methods with model-based methods in deciding and learning how to act through several related scientific tasks. Crucially, empowering an AI agent with the ability to self-deliberate its behaviour and act autonomously carries significant safety risks, which must be guarded by human-guided specifications and oversight, meaning to find a balance of such power with safety. This aspect strongly relates to the agents' trustworthiness theme carried out in WP3.

The work in WP5 is divided into 4 “scientific challenge tasks”, i.e., addressing 4 main scientific challenges in the theme, namely:

- Task 5.1: Extended and multi-facet models of the world dynamics and tasks
- Task 5.2: Integrating data-based methods with model-based methods in deciding and learning how to act

- Task 5.3: Learning for reasoners and planners, and reasoners and planners for learning
- Task 5.4: Monitoring and controlling to make actions AI trustworthy in the real world

We describe the work in each of these challenges in the following. We must stress that most of the research done in this work package is of advanced foundational research in AI with a low technology readiness level (between TRL1 and TRL3). Though, we have a specific task on the potential impact of such foundational studies in future AI systems, also allowing for prospective cross-fertilization with industry to be carried out in WP8, namely:

- Task 5.5: Synergies Industry, Challenges, Roadmap concerning on autonomous actions in AI systems

Finally, we have a specific task dedicated to fostering a scientific community on this theme, namely:

- Task 5.6: Fostering the AI scientific community on the theme of deciding and learning how to act

Next, we describe the work done in each of these tasks separately below. Each challenge task continuously interacts with the others to provide input and receive feedback and challenges to get a closed-loop approach to the research activities. Moreover, synergies with the other scientific work packages WP3 (trustworthy AI), WP4 (Paradigms and Representation), WP6 (multi-agents and social), and WP7 (AutoAI) have been sought.

Task 5.1: Extended and Multi-facet Models of the World Dynamics and Tasks

The challenge of **deliberating the course of action** in partially known environments has been taken by AI mainly within Planning in nondeterministic domains. From the conceptual point of view, the essence of planning is program-synthesis under assumptions (assumptions being the model of the world). However, in Planning typically the task is simply to reach a desired state of affairs. Instead we want to consider agent tasks that can be sophisticated process specifications. For this reason, we adopt logical specification languages developed in Formal Methods. Specifically, we focus on Linear Temporal Logic (LTL), which is the specification formalism most used in Model Checking, except that, when we consider agent tasks, we adopt its finite trace variant LTLf. Observe that much of this work is related to what we may call “Assured Autonomy” i.e., an area of AI concerned with building AI agents that autonomously deliberate how to act under formal guarantees, which merges AI and Formal Methods (FM). Hence, much of the work done here is also related to WP3 (Trustworthy AI) , which indeed includes studying AI systems with formal guarantees.

Beyond nondeterministic domains, partially known environments with multiple agents are the subject of **epistemic planning**, which is relevant for WP6 e.g. through its role in social interaction. The importance of the topic is witnessed by a forthcoming special issue of the AI

Journal with that title. Epistemic planning relies on the resources of epistemic logic in order to enrich the description of planning problems: initial state and goal can be described in a language that is richer than that of classical planning, and the event models of dynamic epistemic logic cater for the description of epistemic action preconditions and effects. First results are published in the AI Journal [Cooper et al. 2021].

Focus on finite traces. The interest in finite traces comes from the consideration that given a task, an intelligent agent should (1) reason, (2) synthesise a course of actions, (3) execute such actions, and (4) be ready for the next task. If the task requires an infinite execution, then the agent would reason only once in its lifetime and then execute the synthesised program forever. While this is perfectly fine if the reasoning is done by the designer as in Formal Methods, it does not make much sense if the reasoning is done by the intelligent agent itself. This observation puts strong emphasis on finite traces in AI. The current advanced techniques for LTLf synthesis use symbolic encoding, with additional various forms of formula/automata decomposition. The very best technique available now has been developed in a paper at ICAPS 2021 [De Giacomo and Favorito, 2021]. We recently are looking at a radically different approach to LTLf synthesis based on forward search: from the formula one builds on the fly an AND-OR graph whose branching factor remains controlled through Knowledge Compilation techniques. This approach shares some similarities with solvers for Fully Observable Non Deterministic (FOND) planning but has the capability of exploring a doubly exponential state space (as needed for LTLf synthesis) instead of a single exponential one. First results are published at IJCAI 2022 [De Giacomo et al., 2022c].

Pure Past LTL. An interesting observation arises when studying logic on finite traces. If we can give the specification in Pure Past Linear Temporal Logic (PPLTL), then because of a property of reverse automata, we can compute the corresponding Deterministic Finite-state Automata (DFA) with at most one exponential blow-up. This result is extremely interesting because it means that one can build symbolic representations of the DFA that are poly-time in the size of the logical specification, see [De Giacomo et al., 2022a]. This observation is at the base of a technique for planning in deterministic and nondeterministic domains for PPLTL formulas with only a polynomial overhead wrt standard planning, hence maintaining essentially the same scalability characteristics of current state-of-the-art planning technologies. We expect this setting to be a sweet spot when handling temporally extended tasks in planning in both deterministic and nondeterministic domains, which has the potential to have a very significant practical impact on the entire Planning in AI area, including on the de facto standard Planning Domain Definition Language (PDDL). See initial work on ArXiv [De Giacomo et al., 2022a].

Non-Markovian Environment Specification. While we focus on tasks specified by finite-trace formalisms, we cannot restrict ourselves to finite traces for environment specifications. Indeed, the environment will not stop working when the agent finishes the task. This calls for finding a well-behaved way of representing the environment's behaviour. One of them formalisms for (nondeterministic) planning domains in Reasoning about Actions and Planning. However, most of this research focused on Markovian behaviour specification, in which the next state of the environment is determined by the previous state, the agent action, and the environment response. Recently this Markovianity has been challenged in several ways, introducing forms of non-Markovian specifications, such as fairness, stability, General Reactivity of Rank 1 (GR(1)), arbitrary safety that remain well-behaved computationally and maintain good scalability properties, possibly the most important results on these within TAILOR are collected in KR 2021 [De Giacomo et al., 2021a], IJCAI 2021 [De Giacomo et al., 2021b].

Data-awareness. The work reported above does not consider unbounded data, which are needed for making our autonomous agent data aware. To do so we need to move from a propositional representation of the state to a first-order one. We use Situation Calculus, developed for Reasoning about Actions in AI as the main target framework for the lifting. We have extended the work on Situation Calculus in several ways, including handling environment's nondeterminism. The key results are presented in the following papers KR 2021 [De Giacomo and Lespérance, 2021], AIJ 2022 [De Giacomo et al., 2021d], IJCAI 2022 [De Giacomo et al., 2022e], IJCAI 2022 [Calvanese et al., 2022]. One important aspect of data-awareness is handling data integration from multiple data sources. To do so, within the work for WP4 we further developed modern description logics, used for formalising conceptual models and ontologies, see IJCAI 2021 [Console et al., 2021], JAIR 2021 [De Giacomo et al., 2021g], FI 2022 [Lembo et al. 2022].

Multiple Environments Models and Best Effort. A crucial observation that the specifications of the environments do not need to be monolithic. That is, we can have a nominal model of the environment and then a model (or more models for that matter) that also includes exceptional behaviours. When the agent synthesises its way of acting for achieving a task it can do so by creating a program that works in both models. However, when the model of the environment starts including too many possible exceptions, a successful way of achieving the task in all models may not exist. To handle these situations, we develop a notion of best-effort programs, i.e., programs that, while may not achieve the task, will do nothing to prevent achieving it. Best-effort programs will in fact achieve the task against a maximal number of (vs. all) possible environment behaviours. In this way, one could have a program that is guaranteed to achieve the task in the nominal model of the environment and is best effort in the models that include too many exceptions. This line of work has been developed through the following key papers KR 2021 [Aminof et al., 2021a], IJCAI 2021 [Aminof et al., 2021b], IJCAI 2022 [Aminof et al., 2022].

Learning and reasoning. Our focus on non-Markovian/Temporal tasks and dynamics leads towards non-traditional forms of reinforcement learning, which merge reasoning and learning. We have developed the notion of Restraining Bolts as an LTLf specification that limits and controls the reinforcement learning process, steering it toward creating strategies to satisfy as much as possible the LTLf specification. We have investigated learning the restraining bolt itself through imitation learning. We have considered forms of restraining bolt specifications that are stemming out of concepts developed in monitoring i.e., defining rewards that depend on prefixes of the desired traces instead of the whole trace itself. We have studied non-Markovian reinforcement learning in which the agent learns non-Markovian dynamics and rewards, still implicitly based on LTLf specifications, developing solutions that simultaneously learn a DFA and solve related Markov Decision Problems (MDPs). Results are already available, in IJCAI 2021 [Ronca and De Giacomo, 2021b], IJCAI 2022 [Ronca et al., 2022b].

Other directions. Agent strategies are deterministic: In any situation they tell the agent what to do to achieve its task. However, there are several cases in which the agent should not commit to any specific strategy. Instead, the agent should choose (and change) the strategy directly during execution. This idea introduces the concepts of nondeterministic strategies, where the agent at every point is given the choice of all actions that would be a step forward in achieving its task. An important result from Discrete Event Control says that there exists a “maximally permissive non deterministic strategy” when we focus on safety specifications. Unfortunately, for LTLf task specifications this result does not apply. However, we have shown that we can still compactly represent all strategies achieving a task specified in LTLf, by two non deterministic strategies, one that allows for deferral and one that does not allow for it, plus a constraint that requires to eventually switch from the deferral one to the non-deferral one. In this way we sort of reconstruct the “maximally permissive non deterministic strategy” of Discrete Event Control in the setting of intelligent agents. This line of work is reported in the following paper, IJCAI 2022 [Zhu and De Giacomo, 2022b] . Related to the idea of not directly considering a single strategy, we have also studied strategies that allow the agent to keep a capability of doing something else while achieving its task, for example, while cleaning the floor of a building the cleaning agent keeps the capability of recharging the battery if it decides to do so. In a sense we give to the agent not only “duties” (the cleaning task to achieve) also “rights” (recharging the battery if needed). Results on this line of work are reported in KR 2022 [Zhu and De Giacomo, 2022a]. We observe that distinguishing agent “duties” and “rights” suggest synergies with intellectual work done within Philosophy and Ethics of AI.

Task 5.2: Integrating Data-based Methods with Model-based Methods in Deciding and Learning How to Act

Task 5.2 is about the use of data-based methods with model-based methods in deciding and learning how to act. This task is led by Hector Geffner, UPF. More precisely, the aim is to study the foundations, techniques, algorithms and tools for integrating data-based learning methods with model-based methods for acting and planning. The integration of learning and reasoning (planning) methods is critical in AI, where current (deep) learning-based methods deliver reactive and opaque boxes ("System 1") that do not generalise properly and make no attempt to understand their environments, while model-based methods rely on models that must be supplied by hand in a suitable language. A key challenge is thus learning symbolic model representations automatically from data, as well as representations of general policies and general problem structure.

We have thus addressed these challenges on three main fronts: learning compact and general representations of environment dynamics (action models or simply models), learning representation of action strategies that generalise to classes of "similar" problems, and learning the subgoal structure of these potentially infinite classes of "similar" problems.

Model learning (Action models, Dynamics). One of the advantages of traditional model-based methods in AI planning is that it provides a crisp notion of "problem similarity" that follows from the languages used for modelling planning problems over discrete state spaces. In fact, two planning problems are deemed "similar" when they are instances of the same planning domain. A planning domain is given by a set of action schemas with lifted preconditions and effects defined in terms of a fixed set of domain predicates. The common structure of a domain given by the action schemas and predicates captures precisely what is common among all "similar" problems, and it is key for obtaining policies and problem subgoals that generalise across all "similar" problems.

A key question is how these first-order symbolic planning domains, captured by action schemas and predicates, can be learned from non-symbolic data. In works reported in ECAI 2020 [B. Bonet, H. Geffner] and in KR 2021 [Rodriguez et al., 2021a], we show how such planning domains can be learned from the state graphs reflecting the structure of the state space of small instances alone. It is well known that a planning instance determines a unique state graph; these approaches address the inverse problem: learning the simplest planning instances over a common domain, that is not known, that generates the observed state graphs. This problem is cast and solved as a combinatorial optimization problem expressed either as a Weighted-MAX SAT problem or as an answer set program. In these works, the states are deemed as "black boxes". More recent work reported at the 2022 ICAPS Workshop on Bridging the Gap Between AI Planning and Reinforcement Learning, considers states expressed as 2-dimensional parsed images, and a similar approach is used to learn planning representations that are grounded on such parsed images [A. Occhipinti, B. Bonet, H. Geffner, 2022].

Learning General Policies. Even with a known compact model, planning is (NP) hard. It is thus natural to ask how to make planning simpler by suitable preprocessing, exploiting the knowledge of a planning domain. In the extreme case, one wants to learn general policies over all possible instances of a given or learned domain. Such policies are called general policies as they are not tied to a particular planning problem nor to a particular state space. A lot of the work in deep learning and deep reinforcement learning has been devoted to learning such policies, in most cases, without learning or using first-order domain representations. The results then have not been robust, nor transparent. In the context of TAILOR, we have thus looked at the problem of learning general policies by leveraging both data and first-order domain models (given or learned, as above). For this, it is crucial to define a domain-independent language for expressing general policies and a way of learning such policies from the data and the models. In work reported at AAAI 2021 [Francès et al., 2021], a novel formulation for expressing and learning such policies is developed. Moreover, the learned general policies are proved to be correct as well. Another paper at AAAI 2021 established correspondences between general policies and the notion of planning width [Bonet and Geffner, 2021], while more recent works at ICAPS 2022 and KR 2022 have shown how similar policies can be learned using graph neural networks [Ståhlberg et al., 2022a,b]. The ICAPS 2022 work was distinguished as the conference Best Paper.

Learning Subgoal Structure (Sketches). The problem of learning subgoal structure appears in both planning and in reinforcement learning where learning in the presence of sparse rewards has been particularly challenging. Some methods developed in the RL setting for addressing this problem involve extra subgoal information in the form of intrinsic rewards, reward machines, or restraining bolts. In most cases, this extra information is provided by hand, in some cases, it is learned automatically, although there has not been a good theory of what the subgoals of a problem should be. Recently, we have developed both a language and a theory for expressing and characterising subgoal structure in the form of sketches. The language of sketches is similar to the language of general policies but the semantics is slightly different. Roughly, a general policy implicitly defines which state transitions (s,s') are "good". The same idea is used in sketches but the state s' does not have to be 1-step away from s; it can be a possible subgoal from s. A sketch thus decomposes problems into subproblems and a sketch is "good" if the resulting subproblems have all bounded widths, and hence can be solved in polynomial time. A paper at AAAI 2021 introduced the language and semantics of sketches [Bonet and Geffner, 2021], and follow up work at KR 2021, showed how to use sketches for taming the complexity of known benchmark domains in planning [Drexler et al., 2021], and a more recent work at ICAPS 2022, how to learn sketches automatically given the common planning domain and some small domain instances [Drexler et al., 2022].

Other. Other relevant works published in this period include a paper at [Bonet and Geffner, JAIR 2020] detailing a planning model, called Qualitative Numerical Planning or QNP, that is suitable for expressing the model abstractions that general policies are aimed to solve [B. Bonet, H. Geffner, 2020] and a paper at ICAPS 2021 [Rodriguez et al., 2021b], developing general algorithms for solving both QNPs and fully-observable, non-deterministic (FOND) planning problems. This work was distinguished as Best Paper at ICAPS 2021. Likewise, our work on learning first-order formulas for characterising dead-end states in planning obtained a Distinguished Paper award at IJCAI 2021 [Stahlberg et al., 2021]. Finally, an overview of our work on learning representations for acting and planning appeared at AAAI 2022 [Geffner, 2022].

Task 5.3: Learning for Reasoners and Planners, and Reasoners and Planners for Learning

Task 5.3 is concerned with the integration of data-based machine learning methods with model-based planning techniques with the aim of developing and studying foundations, techniques, algorithms and tools for integrating learning into reasoners and planners. The challenge is to overcome the fundamental differences of the two approaches: On the one hand, learning approaches are not based on a given model. They are therefore inductive and tend to come without guarantees on correctness or optimality, but the last decade has shown that learning can advance the state of the art significantly in a wide variety of applications. Reasoning approaches, on the other hand, exploit a given model. They deduce a solution from formal (often logic-based) representations and yield a solution with guaranteed soundness or optimality guarantees. The drawback is that reasoning approaches can be somewhat rigid and inflexible, lacking ability to adapt to specific applications.

There are different ways that allow to exploit the potential of learning techniques in a reasoning system without sacrificing correctness and optimality guarantees. Using a provided model as a black-box simulator for a learning algorithm has become popular with the development of Monte-Carlo Tree Search [Kocsis & Szepesvári, 2006] and opened the path for well-informed heuristic and policy networks that guide the search performed by such algorithms [Silver et al., 2016].

A combination of reasoning and learning that maintains the advantages of both is to use learned information in the deduction process of reasoning processes.
We focus here on different ways to combine reasoning and learning that maintain the advantages of both by using learned information in the deduction process of a reasoning process. Progress that exploits this idea has been achieved by learning heuristics for planning, learning to decompose planning tasks into subtasks, dynamic algorithm configuration and planner selection, as well as using learning for plan recognition.

Neural Network Heuristics for Planning Heuristic state space search is among the most successful approaches for automated planning, and improvements with respect to the heuristic (e.g., improved quality or computation time) have been responsible for many advances of the state of art. Heuristic functions are therefore an obvious starting point to introduce machine learning techniques into planning, in particular since planning with a learned heuristic remains sound as long as mild constraints on the safety of the heuristic hold.

Neural networks (NN) are commonly used to learn a function based on labelled input data. In the case of planning heuristics, NNs are trained to approximate a function that takes a state as input and outputs a distance estimate for the state. There are different strategies for the generation of training data, including the computation of shortest paths. As this is only possible in small instances, for states that are sufficiently close to the goal or in domains where data in small instances generalises well to larger ones, [Ferber et al., 2022a] introduce a procedure that makes training data generation scalable through bootstrapping and approximate value iteration, and [Micheli and Valentini, 2021] mitigate this issue by using deep reinforcement learning to train a NN-based heuristic.

Recent NN-based function approximation techniques provide not just the approximated function estimate, but also information on how confident the estimator is in its estimate. In an application like search with NN-based heuristics, it is a natural idea to take this into account, e.g. by falling back to a standard heuristic where confidence is low. [Heller et al., 2022] explore this idea by performing search with multiple open lists that depend on the confidence of heuristic values.

Learning State Space Structure In their analysis of Greedy Best-First Search (GFBS), Heusner, Keller and Helmert [IJCAI 2018, pp. 5269-5273] discovered that there are states in a state space topology that can be used to partition each run of GBFS into phases. Unfortunately, these progress states can only be determined after the search has terminated successfully, and they have therefore been of limited practical interest. Combining learning of states with generalisation among instances in the same domain allows to determine progress states on small instances, compute a formula over description logic features that generalises well over all instances of the domain and exploit the generalised information to scale to large instances. [Ferber et al., 2022b] take a first step in this direction by implementing the sketched algorithm. Their work exploits the learned information by breaking ties in favour of progress states.

Dynamic Algorithm Configuration and Selection Another idea pursued in the context of task 5.3 is to use machine learning to decide which algorithm to apply in which situation or how to configure an algorithm in dependence of the current situation. There are different parts of a heuristic search algorithm that can be selected based on this idea, and in the extreme case the entire planning algorithm is determined dynamically before search has even started.

There have been different ideas that select the right planner based on features of tasks, and the most successful ones are based on NNs [Sievers et al., AAAI 2019; Ma et al., AAAI 2020]. The drawback of the neural network approaches is that the learned models are not interpretable, i.e., it is not clear why a planner is selected and which task features are actually important for the selection. [Ferber and Seipp, 2022] show that complex black-box models are not required to learn strong planner selectors. They train a decision tree which yields equally strong results and allows insights why and in which situation certain planners are selected.

Dynamic algorithm configuration [Biedenkapp et al., ECAI 2020] is a meta-algorithmic approach that uses information about the internal behaviour of an algorithm and information about the instance it is run on to change the configuration of the algorithm during its execution. [Biedenkapp et al., 2022] apply the idea to learn an open list selection policy in a principled and data-driven manner. They not only show that it is possible to learn strong selectors but also gain insights on the advantage of the manually generated open list selection strategy of the LAMA planner [Richter and Westphal, JAIR 2010] .

Learning in Plan Recognition Plan recognition is the task of inferring the actual plan an observed agent (possibly with noisy observations) is performing to achieve a goal. [de A. Santos et al., 2021] and [Rosa Amado et al., 2021] follow two different approaches to tackle the issue: The former encode the problem as a linear program and use reasoning alone to tackle the problem, whereas the latter develop a novel approach to solve both goal and plan recognition tasks simultaneously by combining planning and machine learning techniques to mitigate problems of low and faulty observability. A set of plans is used to train a predictive statistical model of the most likely next states given a set of state observations, and combining these predictive models with landmark heuristics allows to predict the most likely next state given a sequence of observations.

Task 5.4: Monitoring and Controlling to Make Actions AI Trustworthy in the Real World

The main objective of this task is to study foundations, techniques, algorithms and tools for devising and learning meaningful dynamic models that mix human understandable fluents versus human un-understandable features. In particular, we are interested in updating and correcting imperfect models, detecting problems in a model; learning from failures; learning (soft) constraints on the model when the model fails; mixing prior human dynamic knowledge/models with learning from data.

Concerning the goal of mixing prior human dynamic knowledge with learning from data we have developed a method for Online Grounding of Symbolic Planning Domains in Unknown Environments [Lamanna et al., 2022]. In this method an agent can exploit its symbolic knowledge about a planning domain by grounding it in the environment in which it operates. When the environment is initially unknown by the agent, the agent needs to explore it and discover the salient aspects of the environment necessary to reach its goals. Namely, the agent has to discover different aspects of the environments in which it operates. In particular it has to discover the objects present in the environment, the properties of these objects, and their relations, and finally how abstract actions can be successfully executed. We devise a framework that aims to accomplish the aforementioned perspective for an agent that perceives the environment partially and subjectively, through real value sensors (e.g., GPS, and on-board camera) and can operate in the environment through low level actuators (e.g., move forward of 20 cm). We evaluate the proposed architecture in photo-realistic simulated environments, where the sensors are RGB-D on-board camera, GPS and compass, and low level actions include movements, grasping/releasing objects, and manipulating objects. The agent is placed in an unknown environment and asked to find objects of a certain type, place an object on top of another, close or open an object of a certain type. We compare our approach with a state of the art method on object goal navigation based on reinforcement learning, showing better performances. This work builds on top of a general environment for acting, learning, and planning that has been developed in [Lamanna et al., 2021a, Lamanna et al., 2021b].

Task 5.5: Synergies Industry, Challenges, Roadmap Concerning on Autonomous Actions in AI Systems

State of the Art. Research in WP5 aims at empowering the agent with the ability of deliberating on how to act in the world in an autonomous fashion without the direct intervention of humans. Crucially, empowering an AI agent with the ability to self-deliberate its own behaviour carries significant risks of the agent getting out-of-control, hence this ability must be balanced with safety. Assessing safety is essential, and formal verification, model checking and automated synthesis to guarantee safety specifications is central to this effort. This line of research involves several fields of AI, including planning, knowledge representation, logics in AI and probabilistic reasoning as well as verification and automated synthesis in Formal Methods. A current shortcoming in the domain is the gap between theoretical research in planning and industry applications.

Achievements. Several interesting research avenues have been identified by WP5 partners during their research, workshops and meetings. In particular, the following areas are considered important:

- Learning action models (related to WP4);
- Non-Markovian reinforcement learning (e.g. reward machines, temporally extended rewards and dynamics);
- Integrating logic-based reasoning about actions and data-driven learning;
- Learning and acting in robotics (behaviour trees);
- Theory of mind in order to reason about beliefs, capabilities and goals, when deliberating and executing actions (related to WP6);
- Connections and synergies with formal methods;
- Goal reasoning and formation;
- Learning and exploiting automata/goal structure;
- Considering multiple models to handle various levels of contingencies.

Clearly, a main research direction concerns the integration and development of model-based and model-free approaches for learning and planning.

These areas were identified in the TAILOR Strategic Research and Innovation Roadmap (SRIR) as impactful areas for European industry. They in particular concern mobility, production, interacting with humans, fintech, entertainment, and many others. For example, autonomous mobile robot platforms are focusing less on hardware aspects and more on organisation and software, to automate warehouses and logistics. This shift is an opportunity for introducing advanced forms of autonomy based on the kind of work done in WP5. Smart manufacturing could benefit from research in learning and reasoning on how to act by automated program-synthesis and learning how to handle unexpected exceptions. Interaction with humans requires autonomous capability in acting in order not to be too annoying to the humans themselves. FinTech is interested in creating autonomous agents that can act rationally while learning from actual data during operation. Also video games, augmented reality, interactive entertainment is heavily relying on these techniques for improving the interaction and the behaviour of avatars.

As also pointed out in the SRIR, learning and reasoning on how to act is strongly connected with other scientific disciplines outside AI. Acting and planning tools have the potential to boost research and technological development e.g. in formal methods, MDPs, best-effort synthesis, operations research, and cybersecurity. There are also significant connections with the humanities.

Task 5.6: Fostering the AI Scientific Community on the Theme of Deciding and Learning How to Act

State of the Art. The theme of deciding and learning how to act has received considerable attention in the scientific community over the past few years. Papers on this topic now appear regularly at top international AI conferences such as IJCAI, AAAI, ECAI, NeurIPS, and ICML. It also features prominently at specialised conferences such as the *International Conference on Representation Learning (ICLR)*, for example. Moreover, the upcoming AAAI Fall Symposium will also be concerned with this theme in their workshop *Thinking fast and slow and other cognitive theories.*

Achievements. In the context of TAILOR, the following initiatives deserve particular mention: Both as Program co-Chair (2021) and General Chair (2022) of the International Conference on Principles of Knowledge Representation and Reasoning, Gerhard Lakemeyer has fostered a special track on knowledge representation and machine learning that offers a forum for researchers interested in deciding and learning how to act to present their work at the premier international conference on knowledge representation and reasoning. A number of people from the TAILOR network served as members of the program committee or as area chairs. In particular, Luc De Raedt served as co-Chair of the special track in 2021.

Gerhard Lakemeyer, together with Fredrik Heintz and Sheila McIlraith, organised a Dagstuhl seminar on Cognitive Robotics in September 2022. This seminar brought together leading experts in the field of cognitive robotics, knowledge representation, machine learning, and natural language understanding, among others, and included members from the TAILOR network. The themes of the workshop included Cognitive Robotics and Knowledge Representation, Verification of Robotic Systems, Human-Robot Interaction and Ethics, and Planning and Machine Learning. The latter was led by Hector Geffner, member of TAILOR, and featured spotlight talks as well as group discussions. One of the tangible outcomes of the seminar was a collection of challenge problems and a roadmap for future research. These findings will be published in a forthcoming Dagstuhl report.

WP5 Publications

An updated list of all publications related to the WP5 is always available online at https://sites.google.com/diag.uniroma1.it/ict-48-tailor-wp5/papers.

[Abate et al., 2021] Abate, A., Gutierrez, J., Hammond, L., Harrenstein, P.,Kwiatkowska, M., Najib, M., Perelli, G., Steeples, T., and Wooldridge, M. (2021). Rational verification: game-theoretic verification of multi-agent systems. Appl. Intell., 51(9):6569–6584.

[Agostinelli et al., 2021] Agostinelli, S., Bergami, G., Fiorenza, A., Maggi, F., Marrella, A., and Patrizi, F. (2021). Discovering declarative process model behaviour from event logs via model learning. In ICPM, pages 48–55. IEEE.

[Aineto et al., 2021] Aineto, D., Jim´enez, S., and Onaindia, E. (2021). Generalized temporal inference via planning. In KR, pages 22–31.

[Alechina et al., 2022] Alechina, N., De Giacomo, G., Logan, B., and Perelli, G. (2022). Automatic synthesis of dynamic norms for multi-agent systems. In KR.

[Alman et al., 2022] Alman, A., Maggi, F., Montali, M., Patrizi, F., and Rivkin, A. (2022). Multi-model monitoring framework for hybrid process specifications. In CAISE.

[Aminof et al., 2021a] Aminof, B., De Giacomo, G., Lomuscio, A., Murano, A., and Rubin, S. (2021a). Synthesizing Best-Effort strategies under multiple environment specifications. In KR.

[Aminof et al., 2021b] Aminof, B., De Giacomo, G., and Rubin, S. (2021b). Best-effort synthesis: Doing your best is not harder than giving up. In IJCAI.

[Aminof et al., 2022] Aminof, B., De Giacomo, G., Rubin, S., and Zuleger, F. (2022). Beyond strong-cyclic: Doing your best in stochastic environments. In IJCAI, pages 2525–2531.

[Belaid et al., 2022] Belaid, M., Belmecheri, N., Gotlieb, A., Lazaar, N., andSpieker, H. (2022). GEQCA: generic qualitative constraint acquisition. In AAAI.

[Biedenkapp et al., 2022] Biedenkapp, A., Speck, D., Sievers, S., Hutter, F., Lindauer, M., and Seipp, J. (2022). Learning domain-independent policies for open list selection. In PRL Workshop.

[Bonet and Geffner, 2021] Bonet, B. and Geffner, H. (2021). General policies, representations, and planning width. In AAAI.

[Boudou et al., 2021] Boudou, J., Herzig, A., and Troquard, N. (2021). Resource separation in dynamic logic of propositional assignments. J. Log. Algebraic Methods Program., 121:100683.

[Brunori et al., 2021] Brunori, D., Colonnese, S., Cuomo, F., and Iocchi, L. (2021). A reinforcement learning environment for multi-service UAV-enabled wireless systems.

[B ¨uchner et al., 2021] B ¨uchner, C., Keller, T., and Helmert, M. (2021). Exploiting cyclic dependencies in landmark heuristics. In ICAPS.

[Calvanese et al., 2022] Calvanese, D., De Giacomo, G., Montali, M., and Patrizi, F. (2022). Verification and monitoring for first-order LTL with persistence-preserving quantification over finite and infinite traces. In IJCAI, pages 2553–2560.

[Chiariello et al., 2022a] Chiariello, F., Maggi, F., and Patrizi, F. (2022a). ASP-based declarative process mining. In AAAI.

[Chiariello et al., 2022b] Chiariello, F., Maggi, F., and Patrizi, F. (2022b). ASP-based declarative process mining (extended abstract). In Proceedings of the 38th International Conference on Logic Programming (Technical Communications) (ICLP). Electronic Proceedings in Theoretical Computer Science (EPTCS).

[Christen et al., 2022] Christen, R., Eriksson, S., Pommerening, F., and Helmert, M. (2022). Detecting unsolvability based on separating functions. In ICAPS, pages 44–52.

[Cima et al., 2021] Cima, G., Croce, F., and Lenzerini, M. (2021). Query definability and its approximations in ontology-based data management. In CIKM, pages 271–280. ACM.

[Cipollone et al., 2022] Cipollone, R., De Giacomo, G., Favorito, M., Iocchi, L., and Patrizi, F. (2022). Exploiting multiple levels of abstractions in episodic RL via reward shaping. In PRL Workshop.

[Console et al., 2021] Console, M., De Giacomo, G., Lenzerini, M., and Namici, M. (2021). Intensional and extensional views in DL-Lite ontologies. In IJCAI.

[Cooper et al., 2021] Cooper, M. C., Herzig, A., Maffre, F., Maris, F., Perrotin, E., and R ́egnier, P. (2021). A lightweight epistemic logic and its application to planning. Artificial Intelligence, page 103437.

[Corrˆea et al., 2021] Corrˆea, A. B., Franc`es, G., Pommerening, F., and Helmert, M. (2021). Delete-relaxation heuristics for lifted classical planning. In ICAPS.

[Corrˆea et al., 2022] Corrˆea, A. B., Pommerening, F., Helmert, M., and Franc`es, G. (2022). The FF heuristic for lifted classical planning. In AAAI, pages 9716–9723.

[Corrˆea and Seipp, 2022] Corrˆea, A. B. and Seipp, J. (2022). Best-first width search for lifted classical planning. In ICAPS, pages 11–15.

[de A. Santos et al., 2021] de A. Santos, L. R., Meneguzzi, F., Fraga Pereira, R., and Pereira, A. G. (2021). An lp-based approach for goal recognition as planning. In AAAI.

[De Giacomo et al., 2021a] De Giacomo, G., Di Stasio, A., Perelli, G., and Zhu, S. (2021a). Synthesis with mandatory stop actions. In KR.

[De Giacomo et al., 2021b] De Giacomo, G., Di Stasio, A., Tabajara, L. M., Vardi, M., and Zhu, S. (2021b). Finite-trace and generalized-reactivity specifications in temporal synthesis. In IJCAI.

[De Giacomo and Favorito, 2021] De Giacomo, G. and Favorito, M. (2021). Compositional approach to translate ltlf/ldlf into deterministic finite automata. In ICAPS, volume 14.

[De Giacomo et al., 2022a] De Giacomo, G., Favorito, M., and Fuggitti, F. (2022a). Planning for temporally extended goals in pure-past linear temporal logic: A polynomial reduction to standard planning. CoRR, abs/2204.09960.

[De Giacomo et al., 2021c] De Giacomo, G., Favorito, M., Iocchi, L., and Patrizi, F. (2021c). Domain-Independent reward machines for modular integration of planning and learning. In PRL Workshop.

[De Giacomo et al., 2022b] De Giacomo, G., Favorito, M., Leotta, F., Mecella, M., and Silo, L. (2022b). Modeling resilient cyber-physical processes and their composition from digital twins via markov decision processes. CEUR.

[De Giacomo et al., 2022c] De Giacomo, G., Favorito, M., Li, J., Vardi, M. Y., Xiao, S., and Zhu, S. (2022c). Ltlf synthesis as AND-OR graph search: Knowledge compilation at work. In IJCAI, pages 2591–2598.

[De Giacomo et al., 2022d] De Giacomo, G., Favorito, M., Mecella, M., Leotta, F., and Silo, L. (2022d). Digital twins composition via markov decision processes. In ITAL-IA Workshop.

[De Giacomo et al., 2022e] De Giacomo, G., Felli, P., Logan, B. and Patrizi, F., (2022e) Controller Synthesis for Manufacturing Systems in the Situation Calculus. In IJCAI – journal track.

[De Giacomo et al., 2021d] De Giacomo, G., Felli, P., Logan, B., Patrizi, F., and Sardi ̃na, S. (2021d). Situation calculus for controller synthesis in manufacturing systems with first-order state representation. Artificial Intelligence.

[De Giacomo et al., 2021e] De Giacomo, G., Felli, P., Montali, M., and Perelli, G. (2021e). HyperLDLf: a logic for checking properties of finite traces process logs. In IJCAI.

[De Giacomo and Lespérance, 2021] De Giacomo, G. and Lesp ́erance, Y. (2021). The nondeterministic situation calculus. In KR.

[De Giacomo et al., 2021f] De Giacomo, G., Murano, A., Patrizi, F., and Perelli, G. (2021f). Timed trace alignment with metric temporal logic over finite traces. In KR.

[De Giacomo et al., 2021g] De Giacomo, G., Oriol, X., Rosati, R., and Savo, D. F. (2021g). Instance-level update in dl-lite ontologies through first-order rewriting. JAIR, 70:1335–1371.

[De Giacomo and Perelli, 2021] De Giacomo, G. and Perelli, G. (2021). Behavioral QLTL. CoRR, abs/2102.11184.

[Drexler et al., 2021] Drexler D., Seipp, J. and Speck, D. (2021). Subset-saturated transition cost partitioning. In ICAPS, pages 131–139.

[Drexler et al., 2021] Drexler, D., Seipp, J., and Geffner, H. (2021). Expressing and exploiting the common subgoal structure of classical planning domains using sketches. In KR.

[Drexler et al., 2022] Drexler, D., Seipp, J., and Geffner, H. (2022). Learning sketches for decomposing planning problems into subproblems of bounded width. In ICAPS.

[Ferber et al., 2022a] Ferber, P., Geißer, F., Trevizan, F., Helmert, M., and Hoffmann, J. (2022). Neural network heuristic functions for classical planning: Bootstrapping and comparison to other methods. In ICAPS, pages 583–587.

[Ferber et al., 2022b] Patrick Ferber, Liat Cohen, J. S. and Keller, T. (2022). Learning and exploiting progress states in greedy best-first search. In IJCAI, pages 4740–4746.

[Ferber and Seipp, 2022] Ferber, P. and Seipp, J. (2022). Explainable planner selection for classical planning. In AAAI, pages 9741–9749. AAAI Press.

[Fraga Pereira et al., 2021] Fraga Pereira, R., Fuggitti, F., and De Giacomo, G. (2021). Recognizing LTLf/PLTLf goals in fully observable non-deterministic domain models. CoRR, abs/2103.11692.

[Fraga Pereira et al., 2022] Fraga Pereira, R., Messa, F., Grahl Pereira, A., and De Giacomo, G. (2022). Iterative depth-first search for fully observable non-deterministic planning. In ICAPS.

[Francès et al., 2021] Franc`es, G., Bonet, B., and Geffner, H. (2021). Learning general planning policies from small examples without supervision. In AAAI, pages 11801–11808.

[Geffner, 2022] Target Languages (vs. Inductive Biases) for Learning to Act and Plan. Geffner, H. In AAAI, pages 12326–12333, 2022. AAAI Press

[Geisler et al., 2022] Geisler, S., Vidal, M., Cappiello, C., L´oscio, B. F., Gal, A., Jarke, M., Lenzerini, M., Missier, P., Otto, B., Paja, E., Pernici, B., and Rehof, J. (2022). Knowledge-driven data ecosystems toward data transparency. ACM J. Data Inf. Qual., 14(1):3:1–3:12.

[Gutierrez et al., 2021a] Gutierrez, J., Harrenstein, P., Perelli, G., and Wooldridge, M. (2021a). Expressiveness and nash equilibrium in iterated boolean games. ACM Trans. Comput. Logic, 22(2).

[Gutierrez et al., 2021b] Gutierrez, J., Murano, A., Perelli, G., Rubin, S., Steeples, T., and Wooldridge, M. (2021b). Equilibria for games with combined qualitative and quantitative objectives. Acta Informatica, 58(6):585–610.

[Habering et al., 2021] Habering, D., Hofmann, T., and Lakemeyer, G. (2021). Using platform models for a guided explanatory diagnosis generation for mobile robots. In IJCAI, pages 1908–1914. ijcai.org.

[Heller et al., 2022] Heller, D., Ferber, P., Bitterwolf, J., Hein, M., and Hoffmann, J. (2022). Neural network heuristic functions: Taking confidence into account. In SoCS, pages 223–228.

[Helmert et al., 2022] Helmert, M., Sievers, S., Rovner, A., and Corrˆea, A. B. (2022). On the complexity of heuristic synthesis for satisficing classical planning: Potential heuristics and beyond. In ICAPS, pages 124–133.

[Herzig et al., 2022] Herzig, A., Lorini, E., and Perrotin, E. (2022). A computationally grounded logic of 'seeing-to-it-that'. In Raedt, L. D., editor, Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022, pages 2648–2654.

[Herzig et al., 2021] Herzig, A., Maris, F., and Perrotin, E. (2021). A dynamic epistemic logic with finite iteration and parallel composition. In KR, pages 676–680.

[Herzig and Yuste-Ginel, 2021a] Herzig, A. and Yuste-Ginel, A. (2021a). Abstract argumentation with qualitative uncertainty: An analysis in dynamic logic. In Logic and Argumentation - 4th International Conference, CLAR 2021, Hangzhou, China, October 20-22, 2021, Proceedings, volume 13040 of Lecture Notes in Computer Science, pages 190–208. Springer.

[Herzig and Yuste-Ginel, 2021b] Herzig, A. and Yuste-Ginel, A. (2021b). On the epistemic logic of incomplete argumentation frameworks. In KR, pages 681–685.

[Hu and Speck, 2022] Hu, K. and Speck, D. (2022). On bidirectional heuristic search in classical planning: An analysis of bae*. In SoCS, pages 91–99.

[Kin Max Gusm˜ao, 2021] Kin Max Gusm˜ao, Ramon Fraga Pereira, F. M. (2021). Inferring agents preferences as priors for probabilistic goal recognition. In SPARK Workshop (ICAPS).

[Lamanna et al., 2021a] Lamanna, L., Saetti, A., Serafini, L., Gerevini, A. E., and Traverso, P. (2021a). Online learning of action models for PDDL planning. In IJCAI.

[Lamanna et al., 2021b] Lamanna, L., Serafini, L., Gerevini, A. E., Saetti, A., and Traverso, P. (2021b). On-line learning of planning domains from sensor data in PAL: Scaling up to large state spaces. In AAAI.

[Lamanna et al., 2022] Lamanna, L., Serafini, L., Saetti, A., Gerevini, A., and Traverso, P. (2022). Online grounding of symbolic planning domains in unknown environments. In KR.

[Lembo et al. 2022] Lembo, D., Santarelli, V., Savo, D.F. and De Giacomo, G., (2022). Graphol: A Graphical Language for Ontology Modeling Equivalent to OWL 2. In Future Internet, 14(3), p.78.

[Lenzerini et al., 2021] Lenzerini, M., Lepore, L., and Poggi, A. (2021). Meta-modeling and metaquerying in OWL 2 QL. Artificial Intelligence, 292:103432.

[Liu and Lakemeyer, 2021] Liu, D. and Lakemeyer, G. (2021). Reasoning about beliefs and meta-beliefs by regression in an expressive probabilistic action logic. In IJCAI, pages 1951–1958. ijcai.org.

[Max Gusmao et al., 2021] Max Gusmao, K., Fraga Pereira, R., Grahl Pereira, A., and Meneguzzi, F. (2021). Inferring agents preferences as priors for probabilistic goal recognition. CoRR, abs/2102.11791.

[Meneguzzi and Fraga Pereira, 2021] Meneguzzi, F. and Fraga Pereira, R. (2021). A survey on goal recognition as planning. In IJCAI.

[Menguy et al., 2022] Menguy, G., Bardin, S., Lazaar, N., and Gotlieb, A. (2022). Automated program analysis: Revisiting precondition inference through constraint acquisition. In IJCAI.

[Micheli and Valentini, 2021] Micheli, A. and Valentini, A. (2021). Synthesis of search heuristics for temporal planning via reinforcement learning. In AAAI.

[Mittelmann et al., 2021] Mittelmann, M., Herzig, A., and Perrussel, L. (2021). Epistemic reasoning about rationality and bids in auctions. In JELIA, volume 12678 of Lecture Notes in Computer Science, pages 116–130. Springer.

[Mittelmann et al., 2022a] Mittelmann, M., Maubert, B., Murano, A., and Perrussel, L. (2022a). Automated synthesis of mechanisms. In IJCAI, pages 426–432. ijcai.org.

[Mittelmann et al., 2022b] Mittelmann, M., Maubert, B., Murano, A., and Perrussel, L. (2022b). Synthesis of mechanisms with strategy logic (short paper). In ICTCS, CEUR Workshop Proceedings. CEUR-WS.org.

[Oriol et al., 2021] Oriol, X., De Giacomo, G., Estañol, M., and Teniente, E. (2021). Embedding reactive behavior into artifact-centric business process models. FGCS, 117:97–110.

[Parr and Pezzulo, 2021] Parr, T. and Pezzulo, G. (2021). Understanding, explanation, and active inference. Frontiers in Systems Neuroscience, 15.

[Pommerening et al., 2021] Pommerening, F., Keller, T., Halasi, V., Seipp, J., Sievers, S., and Helmert, M. (2021). Dantzig-wolfe decomposition for cost partitioning. In ICAPS.

[Ramon Fraga Pereira, 2021] Ramon Fraga Pereira, Nir Oren, F. M. (2021). Landmark-based approaches for goal recognition as planning. In ICAPS, Journal Track.

[Rodriguez et al., 2021a] Rodriguez, I. D., Bonet, B., Romero, J., and Geffner, H. (2021a). Learning first-order representations for planning from black-box states: New results. In KR.

[Rodriguez et al., 2021b] Rodriguez, I. D., Bonet, B., Sardiña, S., and Geffner, H. (2021b). Flexible FOND planning with explicit fairness assumptions. In ICAPS, volume 31, pages 290–298.

[Ronca and De Giacomo, 2021a] Ronca, A. and De Giacomo, G. (2021a). Efficient PAC reinforcement learning in regular decision processes. In PRL Workshop.

[Ronca and De Giacomo, 2021b] Ronca, A. and De Giacomo, G. (2021b). Efficient PAC reinforcement learning in regular decision processes. In IJCAI.

[Ronca et al., 2022a] Ronca, A., Kaminski, M., Grau, B. C., and Horrocks, I. (2022a). The delay and window size problems in rule-based stream reasoning. Artificial Intelligence, page 103668.

[Ronca et al., 2022b] Ronca, A., Paludo Licks, G., and De Giacomo, G. (2022b). Markov abstractions for PAC reinforcement learning in non-Markov decision processes. In IJCAI.

[Rosa Amado et al., 2021] Rosa Amado, L., Fraga Pereira, R., and Meneguzzi, F. (2021). Combining LSTMs and symbolic approaches for robust plan recognition. In AAMAS.

[Segovia-Aguas et al., 2022a] Segovia-Aguas, J., Celorrio, S. J., Sebastiá, L., and Jonsson, A. (2022a). Scaling-up generalized planning as heuristic search with landmarks. In Proceedings of the International Symposium on Combinatorial Search, volume 15, pages 171–179.

[Segovia-Aguas et al., 2021] Segovia-Aguas, J., Jiménez, S., and Jonsson, A. (2021). Generalized planning as heuristic search. In ICAPS, volume 31, pages 569–577.

[Segovia-Aguas et al., 2022b] Segovia-Aguas, J., Jim ́enez Celorrio, S., and Jonsson, A. (2022b). Computing programs for generalized planning as heuristic search (extended abstract). In IJCAI-22, pages 5334–5338. Sister Conferences Best Papers.

[Seipp, 2021] Seipp, J. (2021). Online saturated cost partitioning for classical planning. In ICAPS.

[Seipp et al., 2021] Seipp, J., Keller, T., and Helmert, M. (2021). Saturated post-hoc optimization for classical planning. In AAAI.

[Sievers et al., 2022] Sievers, S., Gnad, D., and Torralba, A. (2022). Additive pattern databases for decoupled search. In SOCS 2022, pages 180–189. AAAI Press.

[Sievers and Helmert, 2021] Sievers, S. and Helmert, M. (2021). Merge-and-shrink: A compositional theory of transformations of factored transition systems. JAIR, 71:781–883.

[Sievers and Wehrle, 2021] Sievers, S. and Wehrle, M. (2021). On weak stubborn sets in classical planning. In IJCAI, pages 4167–4174.

[Singh et al., 2021] Singh, A., Lipovetzky, N., Ramirez, M., and Segovia-Aguas, J. (2021). Approximate novelty search. In ICAPS, volume 31, pages 349–357.

[Speck and Seipp, 2022] Speck, D. and Seipp, J. (2022). New refinement strategies for cartesian abstractions. In ICAPS, pages 348–352.

[Stahlberg et al., 2022a] Stahlberg, S., Bonet, B., and Geffner, H. (2022a). Learning general optimal policies with graph neural networks: Expressive power, transparency, and limits. In ICAPS, pages 629–637.

[Stahlberg et al., 2022b] Stahlberg, S., Bonet, B., and Geffner, H. (2022b). Learning generalized policies without supervision using GNNs. In KR.

[Stahlberg et al., 2021] Stahlberg, S., Francès, G., and Seipp, J. (2021). Learning generalized unsolvability heuristics for classical planning. In IJCAI, pages 4175–4181.

[Steinmetz et al., 2022] Steinmetz, M., Fiˇser, D., Eniser, H. F., Ferber, P., Gros, T., Heim, P., H ̈oller, D., Schuler, X., W ̈ustholz, V., Christakis, M., and Hoffmann, J. (2022). Debugging a policy: Automatic action-policy testing in AI planning. In ICAPS, pages 353–361.

[Su ́arez-Hern ́andez et al., 2021a] Su ́arez-Hern ́andez, A., Andriella, A., Taranovi ́c, A., Segovia-Aguas, J., Torras, C., and Aleny`a, G. (2021a). Automatic learning of cognitive exercises for socially assistive robotics. In 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), pages 139–146. IEEE.

[Su ́arez-Hern ́andez et al., 2021b] Su ́arez-Hern ́andez, A., Segovia-Aguas, J., Torras, C., and Aleny`a, G. (2021b). Online action recognition. In AAAI.

[Tiger et al., 2021] Tiger, M., Bergstr ̈om, D., Norrstig, A., and Heintz, F. (2021). Enhancing lattice-based motion planning with introspective learning and reasoning. IEEE Robotics and Automation Letters, 6(3):4385–4392.

[Torralba et al., 2021] Torralba, A., Seipp, J., and Sievers, S. (2021). Automatic instance generation for classical planning. In ICAPS.

[Umili et al., 2021] Umili, E., Antonioni, E., Riccio, F., Capobianco, R., Nardi, D., and De Giacomo, G. (2021). Learning a symbolic planning domain through the interaction with continuous environments. In PRL Workshop.

[Viehmann et al., 2021] Viehmann, T., Hofmann, T., and Lakemeyer, G. (2021). Transforming robotic plans with timed automata to solve temporal platform constraints. In IJCAI, pages 2083–2089.

[von Tschammer et al., 2022] von Tschammer, J., Mattm ̈uller, R., and Speck, D. (2022). Loopless top-k planning. In ICAPS, pages 380–384.

[Xiao et al., 2021] Xiao, S., Li, J., Zhu, S., Shi, Y., Pu, G., and Vardi, M. (2021). On-the-fly synthesis for LTL over finite traces. In AAAI.

[Yang et al., 2021] Yang, W.-C., Raskin, J.-F., and De Raedt, L. (2021). Lifted model checking for relational mdps. Machine Learning.

[Zhu and De Giacomo, 2022a] Zhu, S. and De Giacomo, G. (2022a). Act for your duties but maintain your rights. In KR.

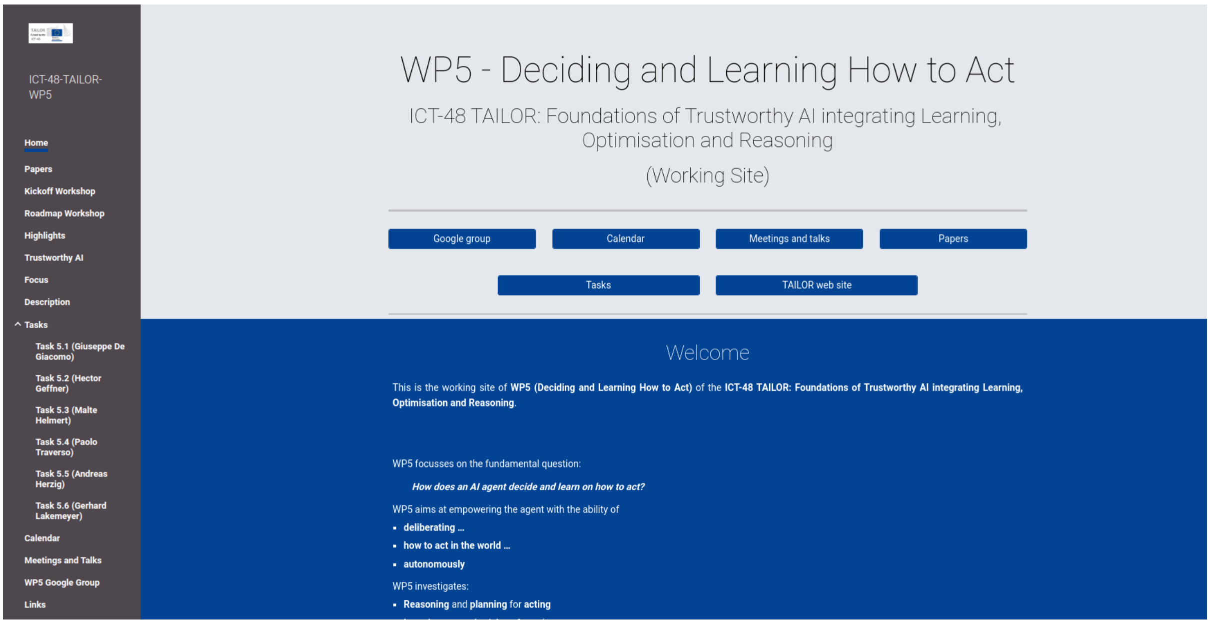
[Zhu and De Giacomo, 2022b] Zhu, S. and De Giacomo, G. (2022b). Synthesis of maximally permissive strategies for ltlf specifications. In IJCAI, pages 2783–2789.

[Zhu et al., 2021] Zhu, S., Tabajara, L. M., Pu, G., and Vardi, M. Y. (2021). On the power of automata minimization in temporal synthesis. In GandALF, volume 346 of EPTCS, pages 117–134.

Appendix 1: Online Website

Due to the unprecedented circumstances brought by the pandemic over the last two years, we decided to coordinate every WP5 activity, meeting, and achievement through an online shared platform identified by a working website available at https://sites.google.com/diag.uniroma1.it/ict-48-tailor-wp5.

Screenshot of the main page.



This working website also features a dedicated forum (https://groups.google.com/g/tailor-wp5-open) for people to exchange ideas, information and collaborate on projects related to WP5.

Appendix 2: Workshops, Open Talks and Courses

Workshops

WP5 Kickoff Workshop (10/12/2020)

Program.
14:45 - 15:00 - Gathering
15:00 - 16:00 - Invited talk by Murray Shanahan (Imperial College London, Deep Mind) - chaired by Hector Geffner - video
16:00 - 16:15 - Break
16:15 - 17:45 - Scientific panels for every task chaired by Kristian Kersting - video

- Task 5.1: Extended and Multi-facet Models of the World Dynamics and Tasks – Giuseppe De Giacomo
- Task 5.2: Integrating Data-based Methods with Model-based Methods in Deciding and Learning How to Act – Hector Geffner
- Task 5.3: Learning for Reasoners and Planners, and Reasoners and Planners for Learning – Malte Helmert
- Task 5.4: Monitoring and Controlling to Make Actions AI Trustworthy in the Real World – Paolo Traverso
- Task 5.5: Synergies Industry, Challenges, Roadmap Concerning on Autonomous Actions in AI Systems – Andreas Herzig
- Task 5.6: Fostering the AI Scientific Community on the Theme of Deciding and Learning How to Act – Gerhard Lakemeyer

17:45 - 18:00 - Break
18:00 - 19:00 - Open discussion on how to organise WP activities (Workshops, micro-projects, site, discussion groups, phd/postdocs managed activities, etc.)

WP5 Roadmap Workshop (20/07/2021)

WP5 has held a workshop, chaired by Andreas Herzig (leader of Task 5.5), on the WP5 contribution to the TAILOR roadmap (WP2).

Program.
14:00 - 15:00 - Invited talk by Sheila McIlraith (University of Toronto, ON, Canada)
15:00 - 15:30 - Presentation of the roadmap by Marc Schoenauer
15:30 - 16:00 - Breakout session
16:00 - 17:00 - Presentation of the breakout session groups and general discussion

Open Talks

WP5 has participated to two TAILOR Open Monthly Meetings, presenting two open scientific talks:

- *Deciding and Learning How to Act* by Giuseppe De Giacomo on March 22, 2022
- *Top-down representation learning for acting and planning* by Hector Geffner on September 6, 2022.

Courses

Artificial Intelligence and Social Intelligence (17/07/2022)

Host Institutions: TAILOR Summer School in Barcelona
Instructor: Andreas Herzig (Centre National de Recherche Scientifique (CNRS) Institut de Recherche en Informatique de Toulouse (IRIT) Univ. Toulouse, France)
Link:
https://www.irit.fr/~Andreas.Herzig/Cours/CSocIntell_Epi/202206_Barca4Tailor_Latex/SocintellEpilogic.pdf

Game-Theoretic Approach to Planning and Synthesis (4/7/2022 – 8/7/2022)

Host Institutions: Sapienza University & ICT-48 TAILOR
Instructors: Giuseppe De Giacomo, Antonio Di Stasio, Giuseppe Perelli, Shufang Zhu (Sapienza University of Rome)
Link: https://whitemech.github.io/courses
Description: This course introduces AI planning and program synthesis for tasks (goals) expressed over finite traces instead of states. Specifically, borrowing from Formal Methods, we will consider tasks and environment specifications expressed in LTL and its finite trace variant LTLf. We will review the main results and algorithmic techniques to handle planning in nondeterministic domains. Then, we will draw connections with verification, and reactive synthesis, together with their game-theoretic solution techniques. The main catch is that working with these logics can be based on devising suitable 2-players games and finding strategies, i.e., plans, to win them. Specifically, we will cover the following topics: Planning in nondeterministic domain, Temporal Logics, LTL, LTLf, Game-theoretic Techniques, Safety Games, Reachability Games, Games for LTL/LTLf objectives, and Reactive Synthesis. This course is partially based on the work carried out in ERC Advanced Grant WhiteMech and EU ICT-48 TAILOR.

Non-Markov Decision Processes and Reinforcement Learning (TBD 7/11/2022 – 21/11/2022)

Host Institutions: Sapienza University with the support of ICT-48 TAILOR and AIDA
Instructors: Giuseppe De Giacomo, Luca Iocchi, Fabio Patrizi, Alessandro Ronca (Sapienza University of Rome)
Guest Lecturers: Roberto Cipollone, Gabriel Paludo Licks, Elena Umili (Sapienza University of Rome)
Link: https://whitemech.github.io/courses
Description: This course is on non-Markov decision processes, where rewards and dynamics can depend on the history of events. This is contrast with Markov Decision Processes, where the dependency is limited to the last state and action. We study how to specify non-Markov reward functions and dynamics functions using Linear Temporal Logic on finite traces. The resulting decision processes are called Regular Decision Processes, and we show how to solve them by extending solution techniques for Markov Decision Processes. Then, we turn to Reinforcement Learning. First, we study the Restraining Bolt, a device that enables an agent to learn a specified non-Markov behaviour while relying on the Markov property. Second, we study how an agent can achieve an optimal behaviour in a non-Markov domain, by learning a finite-state automaton that describes rewards and dynamics. Specifically we will cover the following topics: MDP with Non-Markov Rewards, Non-Markov Dynamics, Regular Decision Processes, Restraining Bolts, Linear Time Logic on finite traces as a reward/dynamics specification language, Reinforcement Learning, Deep Reinforcement Learning, Automata Learning. This course is partially based on the work carried out in ERC Advanced Grant WhiteMech and EU ICT-48 TAILOR.

Appendix 3: Awards

Some work carried out within the WP5 has been recognized with prestigious awards at top-tier international conferences. In particular, the following three papers have been awarded the "Best Paper Award".

- "Learning Generalized Unsolvability Heuristics for Classical Planning". Ståhlberg, S.; Francès, G.; and Seipp. **Best Paper Award at IJCAI 2021**
- "Flexible FOND Planning with Explicit Fairness Assumptions". Rodriguez, I. D; Bonet, B.; Sardiña, S.; and Geffner, H. **Best Paper Award at ICAPS 2021**
- "Learning General Optimal Policies with Graph Neural Networks: Expressive Power, Transparency, and Limits." Ståhlberg, S.; Bonet, B.; and Geffner, H. **Best Paper Award at ICAPS 2022**