# Trustworthiness of AI in Planning and Scheduling: The experience of the TUPLES project


Angelo Gordini [1[0000-0001-8823-9339]], Matteo Pozzi [1[0000-0001-6550-0872]], Michele Lombardi [2[0000-0003-4709-8888]], Thomas Sergeys [3 [0009-0003-3817-4516]], Tias Guns [3 [0000-0002-2156-2155]], Andrea De Cesarei [4[0000-0002-6624-5006]], Sofia Morandini [4[0009-0007-3984-164X]], Sylvie Thiébaux [5[0000-0002-7434-3976]]

[1] OPTIT s.r.l. a socio unico, Bologna, Italy
tuples@optit.net
[2] University of Bologna, Italy
Department of Computer Science and Engineering - DISI
michele.lombardi2@unibo.it
[3] KU Leuven, Belgium
{thomas.sergeys,tias.guns}@kuleuven.be
[4] University of Bologna, Italy
Department of Psychology "Renzo Canestrari"
{andrea.decesarei,sofia.morandini3}@unibo.it
[5] LAAS-CNRS, University of Toulouse, France
Artificial and Natural Intelligence Toulouse Institute (ANITI)
Sylvie.Thiebaux@laas.fr



**Abstract.** Artificial Intelligence and hybrid approaches are increasingly relevant in Decision Science at both the scientific and industrial application levels. While symbolic approaches enjoy well-established techniques to verify their behaviour, data-driven or hybrid approaches can pose a risk in terms of robustness under uncertain conditions, and their behaviour can be less predictable and explainable in unexpected conditions. To address this challenge, the European Commission presented in 2019 the Ethics Guidelines for Trustworthy AI [1], and included in the Horizon Europe 2020-2024 programme a financing of 96 million euros for Trustworthy, Human Centred and Responsible AI. Amongst other initiatives, the TUPLES project [2] was initiated to develop the foundations, approaches and tools needed to achieve transparent, robust, and safe algorithmic solutions in Planning and scheduling (P&S), and increase confidence in these systems to accelerate their adoption. TUPLES is planned to release a toolkit including a set of simulation environments, a self-assessment tool to validate the confidence level of a P&S application, and will launch a competition in late 2024 to engage a wider audience and increase its impact [3]. We outline the challenges, approaches and objectives of the project, focusing on the use cases provided by OPTIT in the application sectors of Energy [4] and Waste management [5]. In doing so, we provide a concrete perspective on how Human Centred AI (with a specific focus on robustness and explainability) requires innovative strategies and approaches that will pave the way for significant improvements in the Decision Support Systems of the coming decades.



**Keywords:** Trustworthy AI, Planning, Scheduling.

**Acknowledgments.** This project has received funding from the European Union's HORIZON-CL4-2021-HUMAN-01 research and innovation programme under grant agreement No. 101070149

# 1 Context

## 1.1 Overview of the approaches in Decision Support Systems

Decision Support Systems (DSS) are information systems that support business or organisational decision-making about problems that may be rapidly changing and that are not easily specified in advance. DSS employ a wide range of approaches in order to simulate, predict, and find the optimal action plan for the examined system, that can broadly classified into:

(i) **Symbolic approaches** i.e. approaches that involve writing a formal description of the system, its constraints, and the resulting "value" of a plan of action (that can be cost, revenue/margin, or other objectives that are relevant for the case). These methods are frequently employed in the field of routing and logistics, scheduling, resource assignment.

(ii) **Data-driven approaches** i.e. approaches that involve training a model on past data related to the behaviour of the system, in order to predict its behaviour in various conditions. These methods are often employed in fields where the human factor is stronger or the domain is less formalised (e.g. marketing, sales, physical systems)

(iii) **Hybrid approaches** that combine both symbolic and data-driven models.

Symbolic approaches are more interpretable, allow the user better control, and typically enable guarantees on the level of constraint satisfaction and other desirable properties of the solution. However, symbolic approaches have trouble dealing with uncertainty, cannot deal with implicit knowledge, typically rely on (possibly inaccurate) hand-crafted models, and have limited scalability.

Data-driven approaches, typically based on machine learning (ML), have complementary properties: they handle uncertainty well and (once learning is finished) can cope with large scale problems. However, they are more opaque, they are less controllable, and typically do not provide guarantees on properties of the solution. In particular there is no guarantee of robustness to environment fluctuations nor a guarantee of safety, i.e. the behaviour induced by a learned policy may drastically change with small input changes and lead into undesirable situations.

Hybrid methods that combine data-driven and symbolic approaches are a promising research field that aims to develop models in a way that benefits from the scalability and modelling power of data-driven approaches, whilst gaining the transparency, robustness and safety of symbolic approaches. This can be achieved, for example, by using ML for predicting parameters (or representing entire parts) of optimization models, or for defining automatically learned heuristics.

Unfortunately, actually realising the advantage of hybrid methods is challenging:

(i) If we are adding ML models to a symbolic system, without additional precautions hybrid DSS may inherit some of the same drawbacks that ML models have. In particular, if a black-box ML model is employed (e.g. a Neural Network), then the resulting DSS will be less interpretable, and its decisions will be ultimately harder to explain; bias in the original dataset may translate into bias in the decisions suggested by the system; vulnerabilities in the ML model may lead to constraint violations.

(ii) If we are adding symbolic components to an ML system, then without additional precautions hybrid DSSs may inherit some of the same drawbacks the symbolic components have, including lack of scalability and high manual costs for model design.

### 1.2 The need for trust

The increasing awareness of both the great benefits and risks of applying artificial intelligence (AI) in real-world settings has given rise to a number of initiatives, in Europe and world-wide, to identify the principles underlying trustworthy AI systems, to understand the technical requirements such systems must satisfy to be trusted, and to provide guidelines for their development.

The European Commission has decided to take a leading role with respect to Human-centred AI, engaging in an internal debate and fostering initiatives that have led to the Ethics guidelines for Trustworthy Artificial Intelligence [1]. However, such guidelines remain very general and abstract, leaving open the novel technical foundations, algorithmic approaches, and tools necessary to implement trustworthy AI systems.

## 2 The TUPLES Project

### 2.1 General information

Amongst other initiatives funded by the Horizon Europe 2020-2024 programme of the European Commission, the TUPLES project [2] was kicked off in October 2022. The overall aim of the project is to investigate suitable combinations of symbolic and data-driven methods in planning and scheduling (P&S) that accentuate their strengths while compensating for their weaknesses. The hypothesis of the project, backed up by the above cited EU guidelines, is that any trustworthy AI system, and in particular any system for P&S, must exhibit the properties listed in Table 1 in order to be trusted.

**Table 1.** Trustworthiness properties of P&S algorithms.

| Property | Definition |
|---|---|
| Transparency | Being able to explain how it came to a plan, and why particular decisions were favoured over alternative courses of action. To trust DSS, users need to understand their capabilities and limitations. AI-based DSSs should be able to explain and justify their decisions in a way that is clear and comprehensible. |
| Robustness | Not responding dramatically to irrelevant changes in the input or small fluctuations in the environment. |

| | |
|---|---|
| | DSS decisions need to be robust in terms of being insensitive to irrelevant environment changes, while remaining sensitive to relevant ones and sufficiently general to cover relevant situations. |
| Safety | Avoid recommending plans that could lead to undesirable situations, especially when human lives are involved and there is a risk of injuries or death in case of misbehaviour of the system. |
| Scalability | Being able to tackle significant problems from the real world, not just laboratory-size ones. |

These use cases shown in Table 2 are being specifically investigated by TUPLES.

**Table 2.** Use cases investigated by the TUPLES project.

| Use case | Provided by |
|---|---|
| Scheduling workflow in a manufacturing environment | AIRBUS |
| Virtual assistant for single-pilot aircraft operation | AIRBUS |
| Squad management for professional football clubs | SCISPORTS |
| Smart energy systems supply optimisation | OPTIT |
| Waste collection optimisation | OPTIT |

The research groups involved in the project are:

(i) **ANITI**, Artificial and Natural Intelligence Toulouse Institute.

(ii) **AIRBUS**, Europe's main commercial aircraft manufacturer.

(iii) **KU Leuven**, renowned for its pioneering work on hybrid AI, in particular probabilistic programing, statistical relational learning, predictive learning, decision-focussed prediction, and verification of ML systems.

(iv) **Saarland University**, one of the leading institutions for Computer Science (CS) and automated planning, including planning under uncertainty, and verification and explanation of plans.

(v) **University of Bologna**, one of Europe's top laboratories for optimisation methods combining heterogeneous techniques and their applications to routing, scheduling and logistics. The Department of CS and Engineering has also contributed to the European guidelines for Trustworthy AI [1], and the Department of Psychology is conducting research on risk, safety, acceptance of technology and human-machine interfaces [6,7].

(vi) **Czech Technical University**, the top technical university in the EU, with over 20.000 students and over 3.000 academics and scientists.

(vii) **SciSports**, a spinoff of the University of Twente, and a leading provider of football data intelligence for professional football organisations, football players, media and entertainment.

(viii) **OPTIT**, a spinoff of the University of Bologna Operations Research team, and a leading company for the development of innovative Decision Support Systems, integrating forecasting, data analytics, simulation and optimisation.

## 2.2 Research approach

The research activities of the TUPLES project can be classified in three main areas:

(i) New techniques aimed at improving transparency, robustness, safety and scalability of AI planning/scheduling systems through the combined use of data driven and symbolic methods

(ii) New methods to verify and test DSS for planning and scheduling, with a strong emphasis on approaches featuring Machine Learning components

(iii) New methods for explaining the output of DSS for planning and scheduling, including techniques to support Human in the Loop/Human in Command schemes and connections to ethical aspects

The current industry standard to combine machine learning and planning/scheduling is to treat the two separately and in isolation: historic data and related features are collected; and demand, traffic flows, weather etc. are forecasted using machine learning methods and standard loss functions like mean squared error or mean absolute percentage error. This ML step is followed by using standard scheduling and planning software on the point predictions, as if they are given as facts.

An increasing number of works [10,11,12,13,14,15] show that hybrid learning and scheduling methods, including end-to-end learning methods, can improve on this practice, through ‘in the loop’ learning: they use the actions and decisions of the DSS to determine the outcome and loss values, to guide the training by the task-specific errors that truly matter. However, this raises new issues in terms of robustness, scalability and feedback mechanisms that have to be addressed by further research.

Explainable AI (XAI) is another pillar of trustworthy AI. Model agnostic approaches represent one of the most visible XAI approaches, yet do not offer guarantees of rigour, and this can have disastrous consequences in high-risk or safety-critical situations. Some authors have proposed intrinsic explainability, i.e. the model serves as the explanation, yet intrinsic explainability may propose explanations arbitrarily larger than minimal explanations. Recent work has investigated logic-based methods for explainability, with a number of key contributions by members of the consortium.

The overall goal of the TUPLES project is to devise novel solutions for explaining P&S algorithms.

## 2.3 Expected outcomes of the project

One of the main practical results of the TUPLES project will be the TUPLES Toolkit that will be publicly released in late 2024, and will be composed of:

(i) The TUPLES lab, a set of laboratory/simulation environments designed for controlled experiments, and which will serve as a common development layer for the techniques devised in the project;

(ii) One or more extensions to the scikit-decide environment from Airbus, which provides a framework for applying sequential decision-making algorithms (planning, scheduling and reinforcement learning) to multiple application domains.

(iii) The TUPLES self-assessment tool, an easy-to-use diagnostic survey, whose design will be based on the experience matured in the course of the project, that any solution provider may adopt to evaluate the coherence of a specific approach with respect to the key EU guidelines on trustworthy, reliable, robust and safe AI. It will be implemented as a web survey capitalising on the know-how developed during the project to transform the high level EU guidelines on trustworthiness into pragmatic questions and evaluation metrics. This will help designers of current and prospective DSS to develop a systemic approach to trustworthy AI that can be integrated in a company's design approach.

The TUPLES consortium will launch a competition in late 2024, based on one of the project use cases shown in Table 2, and leveraging the TUPLES Toolkit.

# 3 Smart energy systems supply optimisation

## 3.1 Problem description and issues

Smart energy systems in the District Energy industry are becoming increasingly complex, managing a diversified energy mix, several technologies with radically different characteristics, aiming to maximise not only an economic, but also environmental sets of objectives in a volatile environment both on the demand (heat/electricity) and market (energy prices) side. Endothermal engines and gas boilers are complemented by heat pumps, heat storages, electric boilers, batteries, as well as electric or absorption chillers, and renewable/waste heat sources.

OPTIT has a long lasting experience developing DSS to improve management and planning of complex Heat & Power plants for leading Italian and EU utilities. Based on this experience, a resolution approach was developed that spans from demand forecasting to long (one year) via short-term (next day) planning optimization, evolving further towards intraday (same-day) energy trading. This includes use of diverse Machine Learning techniques for predictive modelling, Mixed Integer Linear Programming (MILP) models for planning modules, supported by a series of math-heuristic approaches. Yet, these components currently adopt a deterministic approach, not fully adequate to respond to the real market challenges.

## 3.2 Example case under investigation

For the TUPLES project, we decided to focus mainly on the short term management of generation plants, also known as Unit commitment problem, comprising a set of generating units (each with its own characteristics and technical limitations) that have to be operated on a relatively short time horizon (e.g. a few days) in order to meet the overall amount of heat (or cooling, or electricity for internal use) requested to the plant, then the excess electrical energy can be sold to the market at a price, maximising the overall EBITDA of the plant (i.e. the earnings before interest, taxes, depreciation, and amortisation), derived by subtracting from revenues all costs of the operating business, but not decline in asset value, cost of borrowing, lease expenses, and obligations to governments.

Plant managers are responsible for managing all the activities of one or more sets of generation assets, ensuring that the energy demand is satisfied at any given time while keeping the economics of the system sustainable.

The problem is currently managed as a deterministic MILP problem, thus belonging to the symbolic methods, but the business presents several elements of uncertainty, such as heating or cooling demand, electric market prices, and production unit unavailability. Solutions need to be robust against these uncertainties.

For the TUPLES project, OPTIT provided a fake generation plant based on the mix-up of two existing plants, to provide a generalisation of most real plants, together with one year of energy demand and electricity market prices with hourly detail.

The plant is internally described as a graph, connecting different nodes with their respective capacity to provide, transform, or consume energy (see Fig. 1). The files describing the composition and characteristics of the generation plant used for the tests are available at [21].

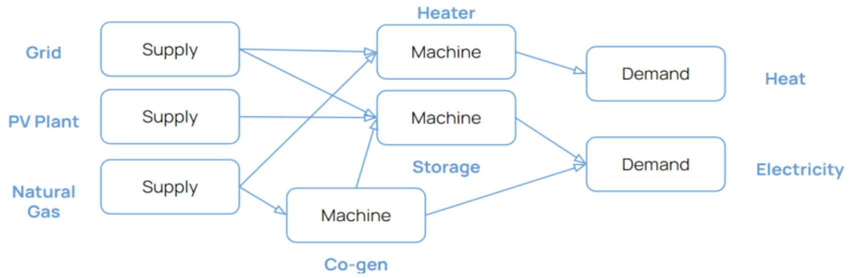


**Fig. 1.** Graph depicting the reference plant for the use case.

### 3.3 Trustworthiness metrics and objectives

For the use case provided, we identified these trustworthiness metrics:

(i) **Robustness**: the planner, given a variation of 10% in the energy demand, should change the set of operating assets in less than 10% of the cases, as measured on a representative sample of historical data.

(ii) **Explainability**: even if the capability to provide explainable and convincing solutions to the end user is a key factor for the DSS adoption, at a first investigation this can be addressed by simulating counterfactuals and showing the user that the alternative plans are worse or not feasible..

### 3.4 Ongoing work

The Energy use case has been the first implementation of the TUPLES Lab, involving a simulation and data generation environment capable of:

(i) Generating predictions for configurable experimental setups;
(ii) Simulating the effect of both epistemic and aleatoric uncertainty;
(iii) Evaluating the quality of a plan under uncertainty via a Monte-Carlo approach with simulated optimal recourse actions (i.e. plan adjustments performed by the plant operator to account for observed uncertainty)

The environment is a necessary ingredient for evaluating robustness in realistic conditions, which always involve some degree of uncertainty. An instance of this first TUPLES Lab environment consists in:

(i) A description of the plant;

(ii) A set of price and demand predictions;
(iii) A variance model, specifically a Gaussian Process trained on historical data that specifies how prices and demands can deviate from the predictions.

The plan simulator operates in an multi-step fashion: at every step the variance model is first sampled to obtain realisation for the uncertain quantities (prices and demands), then the recourse actions of the plant operator is computed by solving a simplified, single-time-ahead, optimization problem (see Fig. 2). Once the simulation is over, the environment enables computing metrics to evaluate the robustness of the plan.

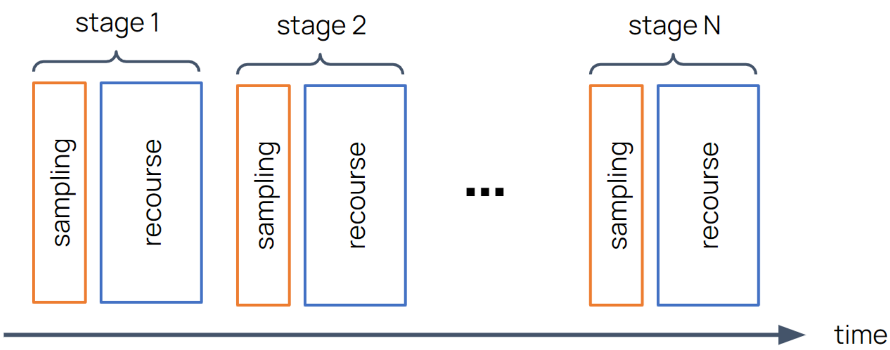


**Fig. 2.** Multi-step simulation for evaluating robustness.

The simulator also provides the starting point for the design of a hybrid method that aims at improving the robustness of the current solver. The approach will be based on a generalisation of Decision Focused Learning (DFL) that relies on black-box differentiation to enable task-specific learning of ML models used to predict parameters of optimization problems. The technique is a result of research conducted prior to and during the TUPLES project, which led to the UNIFY method [8] and a Score Function Gradient Estimation method capable of addressing two-stage stochastic optimization problems [9]. In detail, the approach builds over a deterministic optimization formulation, which can be viewed as a function:

$$z^*(x, y) \equiv \underset{z'}{\operatorname{argmin}}\{f(x, y, z) \mid z \in C(x, y)\}$$

where x represents a vector of parameters of the optimization problem that are fully known at solution time and y represents a vector of problem parameters that are subject to uncertainty, but treated as deterministic by the solver.

The C(x, y) function defines the set of feasible decisions z, based on the problem parameters, while f(x, y, z) represents the decision cost, again estimated by assuming determinism. For the considered use case, the OPTIT solver provides a viable implementation for this component. Our method then attempts to account for the effect of uncertainty by using a ML model to provide a vector of estimated parameters $\hat{y}$ that can indirectly account for the effect of uncertainty on the decision cost.

This is achieved by training the model for minimal (expected) decision cost, rather than for minimal accuracy. Formally, the training problem is formulated as:

$$\underset{\theta}{\operatorname{argmin}}\, \mathrm{E}_{x,y\in P(XY)}\left[F(x,y,z^*(x,\hat{y}))\right]$$
$$\text{with: } \hat{y} = h(x;\theta)$$

where *P(XY)* represents the distribution of all problem parameters, and *h* is a differentiable Machine Learning model with parameter vector $\theta$ (typically a Neural Network). The *F(x, y, z)* function evaluates the decision cost, including the effect of resource actions that can be performed at execution time *once the uncertain parameters become known*. Intuitively, based on the values of the known parameters x, the ML models provides an estimated parameter vector $\hat{y}$; this is employed by the solver to compute a decision vector $z^*$, whose cost is then evaluated by taking into account the corrections that (e.g.) a plant operator can make to ensure that the customer demands are met. Compared to more traditional approaches, where the ML model is trained to make accurate predictions, the proposed method incentivizes the model to yield vectors $\hat{y}$ can deviate from the estimated average, as long as that leads to a better cost.

The key technical difficult is that, while the value of $F(x, y, z^*(x, \hat{y}))$ can be evaluated via simulation, the function lacks a closed-form differentiable formula. The methods developed during the project address this issue by leveraging black-box optimization methods originating in the Reinforcement Learning (RL) literature.

The result is a very versatile hybrid approach that combines RL and Constrained Optimization, using the former to handle uncertainty and the latter to enforce constraint satisfaction. The approach admits a remarkably simple implementation if standard RL libraries are used. The technique can be however slow to converge, which is a significant drawback for practical problems and one of the main challenges to be addressed in the forthcoming months of research.

To facilitate the implementation of these techniques to the use case and specifically the training of the Machine Learning model, OPTIT set up a REST service for the deterministic solver for the reference plant to be remotely invoked by the research teams. The service employs a simplified API data exchange based on two JSON files: the input file, that has to be sent as the body of the POST, contains the demand and prices for one or more days; the output file, that is the body of the response, contains the setpoints and flows calculated by the deterministic solver. A set of example JSON input files is available at [22].

## 4 Waste collection optimisation

### 4.1 Problem description and issues

Recyclable waste is collected and transported to treatment plants every day, employing a large amount of resources with significant impacts on the environment. OPTIT has developed a suite of tools to support strategic and operational service design, based on advanced routing approaches coupled with clustering techniques and several data management and analytics functionalities.

The objective of such methodologies is to enable the Utility planner to increase service levels and operational efficiency, while reducing capital expenditures (vehicles, resources), operational costs (fuel, time shifts) with a positive impact on the environment ($CO_2$, pollutants). The decision maker has to choose how many trucks to use, which routes should be performed by each truck, and which bins have to be collected by which truck. The objective is to keep costs low and provide high quality and fair service, all while ensuring acceptable working conditions for the drivers, and minimising impacts (e.g. traffic congestions) on the citizens. Truck capacities and work shift duration cannot be exceeded, but travel times and amount of waste to be collected can be difficult to estimate in an accurate and consistent way.

In symbolic DSS approaches for the waste collection, the user has to manually specify the constraints and decision variables: this implies that they are interpretable, at least to some degree, and that desired behaviours can be forced by simply adding constraints, so the system would return only solutions that satisfy all specified constraints. On the other hand, any constraint that is not explicitly known (e.g. best practices, or long-standing agreements on work shift) would be completely ignored; some constraints may be very complex to formulate; and constraint satisfaction guarantees can still fail if quantitative estimates are inaccurate (e.g. bin levels).

In data-driven approaches, implicit constraints are naturally taken into account and complex constraints may be easier to implement. However, it is usually difficult to explain the reason for the actions suggested by such algorithms, and their development requires either a large amount of data or to set up a simulation environment.

### 4.2 Example case under investigation

For the TUPLES project, OPTIT provided an artificial door-to-door waste collection case, based on the road map of a real Italian city (10863 road arcs). We also placed randomly throughout the city 3516 collection points to be visited twice a week, one truck depot, and one waste treatment plant (see Fig. 3).

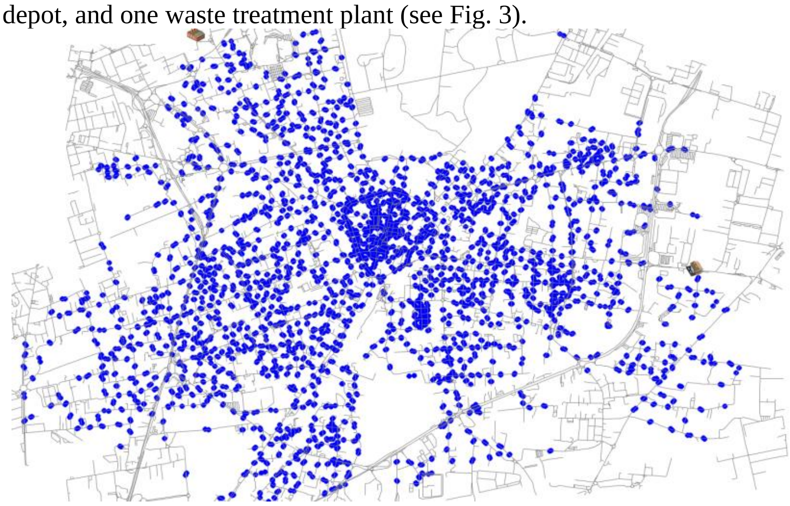

**Fig. 3.** Geographical visualisation of the reference instance for the Waste use case.

Given a maximum truck capacity, a maximum fleet size, and a maximum duration of a shift, the algorithm should provide to each driver a navigation map of the points to be visited twice a week. The reference instance is available at [23] and an example solution is available at [24], both in forms of ESRI shapefiles.

The problem is multi-objective, because different metrics have to be taken into account, and not all of the metrics (or at least the respective weight) have been fully investigated and validated. The metrics involved in the problem include:

(i) Total number of trucks and shifts;
(ii) Total distance of trips;
(iii) Total duration of trips;
(iv) Compactness of trips;
(v) Overlapping of areas covered by trips;
(vi) Fairness of duration between the trips.

### 4.3 Trustworthiness metrics and objectives

Issues such as loss of solution compactness or the service of contiguous bins by different routes may reduce acceptability of the solutions provided. It is therefore crucial to explain the advantages of such solutions over different ones that can be produced manually by a human operator. For the use case provided, we identified these trustworthiness metrics:

(i) **Robustness**: whilst being important for the solutions to be robust with respect to uncertainties on boundary conditions, such as quantity of collected waste, service-time duration, and travel-time duration, this can be addressed by adjusting the constraints keeping a security margin or simulating the proposed plan with different parameters, so it will not be further investigated by the research teams.
(ii) **Explainability**: the Decision Support System should be able to explain why a given collection point (or part of a trip) has been assigned to that trip instead of another.

### 4.4 Ongoing work

The Psychology group is evaluating how to integrate human knowledge and experience in the model. This means that expert human knowledge can be used to constrain possible solutions or to shape the system’s explanations depending on the characteristics of the user. This can be done by taking into account insights based on real-world experiences and observations, informal constraints, and problem-solving strategies that may not be observable through data. Thus, human-centred feedback will be aimed at adapting AI capabilities to the practical realities, challenges and individual characteristics of planners and schedulers. The goal is to define a concrete (numerical) way to translate the "beauty" metrics, i.e. how to define in a quantitative way what is a beautiful (and convenient) trip for the users.

The integration of human-centred feedback underscores the essential role of trust in AI-human interactions [7]. Understanding the link between user trust and the system's ability to integrate human knowledge enhances the model alignment with user needs and expectations. Trust can be measured, and it depends on both an initial assessment and a further dynamic interaction with the system. Some of the factors that modulate trust in AI systems are the ability to adapt to user needs, the presence/clarity of the explanations provided, and the consistency in the results over time [16]. Furthermore, given that human users and the AI system collaborate in the execution of a task (shared agency), it is important to consider the agreement between the decisions made by the AI system and by humans [17,18].

In an ongoing study, we are investigating the relative impact that human-AI agreement and outcome quality have on trust in the AI system, and the moderating role of explanations. For example, it is possible that, in the case of positive outcomes of an AI decision, the degree of similarity to the decision that a human user would have made is perceived as relatively unimportant; however, if an AI decision differs significantly from human decisions and leads to negative outcomes, then trust in the AI system might drop significantly. The effects of agreement and outcome quality on trust may be modulated by the presence and type of explanations that the system gives to motivate its decisions. Knowing the respective roles of explanations, agreement, and outcome quality on user trust may help choosing between different solutions, when the numeric value of two different solutions is similar, but one of them is significantly at odds with human choices.

Regarding the explanations, we are also looking into techniques from explainable constraint solving [19]. The desired explanations are contrastive explanations that, given a base solution and a foil (a part of the assignment which should be different), computes the minimal changes to make the foil true. Existing optimization techniques are applicable here, though there are design dimensions in terms of defining the minimal changes, e.g. the number of changes to the solution, the magnitude of them, or changes/magnitude in terms of pre-defined sub-objectives/KPIs. Furthermore the contrastive explanation may lead to further contrastive questions (e.g. different what-if analysis) as well as modifications to the problem definition. This poses the challenge of an interactive and incremental approach to solving the problem, including the computation of multiple alternative solutions.

In order to provide more acceptable solutions in the first place, alternative techniques are also being experimented, such as Constrained clustering [20]. Provided that this is a multi-objective problem, we are also considering presenting the user multiple possible non-dominated solutions (see Fig. 4), allowing the users to choose which is preferred, in order to improve the search of solutions taking into account a more validated version of the user preferences.

| | nr of tours | total duration | total km | min duration | max duration | duration img | min volume | max volume | volume img |
|---|---|---|---|---|---|---|---|---|---|
| | 76 | 15:10:41:58 | 15524 | 03:36:32 | 05:31:43 | | 119 | 349 | |
| | 76 | 15:09:14:00 | 15377 | 03:25:03 | 05:32:50 | | 104 | 349 | |
| | 77 | 15:12:20:45 | 15658 | 03:20:10 | 05:22:25 | | 161 | 349 | |
| | 83 | 15:23:59:24 | 16616 | 03:04:18 | 04:59:33 | | 88 | 349 | |

**Fig. 4.** Visualisation of non-dominated solutions of the problem.

## 5 Conclusions

The TUPLES project, funded by the EU within the Horizon Europe 2020-2024 programme, aims at developing the foundations, approaches and tools needed to achieve transparent, robust, and safe algorithmic solutions in Planning and scheduling (P&S), and increasing confidence in these systems to accelerate their adoption.

Human-Centred AI represents one of the key factors to foster improvement and adoption of Decision Support Systems of the coming decades and TUPLES represents a significant action to integrate theoretical approaches and practice.

Use cases have been identified and shared among project's partners, trustworthiness metrics and objectives have been identified for each case, and the Energy and Waste use cases provided by OPTIT have been examined in detail in the paper. Research groups are investigating the application of different techniques (Decision focused learning etc.) to improve and possibly guarantee the trustworthiness (specifically the robustness and explainability) of the employed planning algorithms..

TUPLES is planned to release a toolkit including a set of simulation environments, a self-assessment tool to validate the confidence level of a P&S application, and will launch a competition in late 2024 to engage a wider audience and increase its impact.

## References

1. Ethics guidelines for Trustworthy AI, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
2. TUPLES description, Trustworthy Planning and Scheduling with Learning and Explanations, https://cordis.europa.eu/project/id/101070149
3. TUPLES website, Trustworthy Planning and Scheduling with Learning and Explanations, https://tuples.ai/
4. OPTIT website, Energy Management DSS, https://www.optit.net/en/solutions/energy/
5. OPTIT website, Waste Management DSS, https://www.optit.net/en/solutions/waste/
6. Morandini, S., Fraboni, F., Balatti, E., Hackmann, A., Brendel, H., Puzzo, G., Volpi, L., Giusino, D., Angelis, M., Pietrantoni, L. (2023). Assessing the Transparency and Explainability of AI Algorithms in Planning and Scheduling tools: A Review of the Literature. In: Tareq Ahram and Redha Taiar (eds) Human Interaction & Emerging Technologies (IHIET 2023): Artificial Intelligence & Future Applications. AHFE (2023) International Conference. AHFE Open Access, vol 111. AHFE International, USA. http://doi.org/10.54941/ahfe1004068
7. Morandini, S., Fraboni, F., Puzzo, G., Giusino, D., Volpi, L., Brendel, H., Pietrantoni, L. (2023). Examining the Nexus between Explainability of AI Systems and User's Trust: A cPreliminary Scoping Review. *Proceedings http://ceur-ws. org ISSN*, *1613*, 0073.
8. Silvestri M., De Filippo A., Lombardi M., Milano M. (2022). UNIFY: a Unified Policy Designing Framework for Solving Constrained Optimization Problems with Machine Learning. https://doi.org/10.48550/arXiv.2210.14030
9. Silvestri M., De Filippo A., Lombardi M., Berden S., Mahmutogulları A.I., Guns T., Mandi J., Mulamba M. (2023). Score Function Gradient Estimation to Widen the Applicability of Decision-Focused Learning. International Conference on Machine Learning (ICML), 2023. https://doi.org/10.48550/arXiv.2307.05213
10. Elmachtoub, Adam N., Grigas P. (2022) Smart "predict, then optimize". Management Science 68, no. 1 (2022): 9-26.

11. Wilder, B., Dilkina, B., Tambe M. (2019). Melding the Data-Decisions Pipeline: Decision-Focused Learning for Combinatorial Optimization. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 1658-1665. https://doi.org/10.1609/aaai.v33i01.33011658
12. Vlastelica M. et al., Differentiation of blackbox combinatorial solvers. (2019) https://arxiv.org/abs/1912.02175
13. Mandi, Jayanta, Stuckey P. J., and Guns T. Smart predict-and-optimize for hard combinatorial optimization problems. In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 02, pp. 1603-1610. 2020.
14. Mandi, Jayanta, Kotary J, Berden S, Mulamba M, Bucarey V, Guns T, Fioretto F,. "Decision-focused learning: Foundations, state of the art, benchmark and future opportunities." arXiv preprint arXiv:2307.13565 (2023).
15. Kotary J, Fioretto F, Van Hentenryck P, Wilder B. "End-to-End Constrained Optimization Learning: A Survey." International Joint Conference on Artificial Intelligence. (2021.)
16. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1-38.
17. Narayanan S, Guanghui Yu, Chien-Ju Ho, Ming Yin. (2023). How does Value Similarity affect Human Reliance in AI-Assisted Ethical Decision Making?. In AAAI/ACM Conference on AI, Ethics, and Society (AIES '23), August 08–10, 2023 ,Montréal, QC, Canada. ACM, NewYork, NY, USA.
18. Mehrotra S, Jonker C.M., Tielman M. L (2021). More Similar Values, More Trust? - The Effect of Value Similarity on Trust in Human-Agent Interaction. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21), May 19–21, 2021, Virtual Event, USA. ACM, New York, NY, USA,s.
19. Bleukx I, Guns T., Tsouros D. (2024). CPMpy/XCP-explain: Recorded version (v1.0). Zenodo. https://doi.org/10.5281/zenodo.10694140
20. Wagstaff K, Cardie C, Rogers S, Schroedl S. (2001) Constrained K-means Clustering with Background Knowledge. In Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, June 28 - July 1, 2001. https://www.cs.cmu.edu/~dgovinda/pdf/icml-2001.pdf
21. Block-oriented description of a fictional CHP generation plant for unit commitment testing, https://doi.org/10.5281/zenodo.10875998
22. Dataset for testing unit commitment models of CHCP plants, https://doi.org/10.5281/zenodo.10875471
23. Example instance of Waste collection services for testing routing algorithms, https://doi.org/10.5281/zenodo.10875680
24. Example solution of Waste collection routing for benchmarking routing algorithms, https://doi.org/10.5281/zenodo.10875837