UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat d'Informàtica de Barcelona

FIB

SEARCH & ANALYSIS OF NEW HEURISTICS FOR SOLVING NP-HARD PROBLEMS WITH DEEP REINFORCEMENT LEARNING

TOMÀS OSARTE SEGURA

Thesis supervisor: SERGIO ÁLVAREZ NAPAGAO (Department of Computer Science)

Degree: Bachelor's Degree in Informatics Engineering (Computing)

Bachelor's thesis

Facultat d'Informàtica de Barcelona (FIB)

Universitat Politècnica de Catalunya (UPC) - BarcelonaTech

Contents

List of Figures

List of Tables

1 Introduction and Contextualization

This type A graduation thesis is situated in the scope of the *Facultat d'Informàtica de Barcelona* (FIB) and pretends to develop a new heuristic with the Deep Reinforcement Learning framework with the ambition to exceed the state of the art.

In the field of computational complexity theory, a problem is called NP-hard if for all other problems which can be solved in non-deterministic polynomial-time, there exists a polynomial-time reduction to the problem, as it is explained in [1]. This type of problems have been recurrently studied in the field of computation and more recently in the last decade, a new approach to find some heuristics to solve them in a short amount of time have been developed using the framework of reinforcement learning. In this thesis, we will address the subset of combinatorial optimization problems known to be NP-hard.

1.1 Basic concept definition

In order to proceed with more dense explanations, there are some basic concepts that should be defined and properly explained.

1.1.1 NP-Hard Problems

NP-hard problems are computational challenges that are notoriously difficult to solve efficiently. They encompass a wide range of optimization, scheduling, and decision problems across various domains. Examples include the Traveling Salesman Problem, the Knapsack Problem, and the Boolean Satisfiability Problem.

These problems are relevant because they capture the complexity of many real-world scenarios and are fundamental in computer science and related fields. While finding exact solutions to NP-hard problems is often impractical, research into approximation algorithms and heuristic methods continues to drive innovation in algorithm design and optimization techniques. Overall, NP-hard problems pose significant computational challenges and inspire ongoing research efforts to develop more efficient solution strategies. For further details, readers are encouraged to consult [1, 2, 3].

1.1.2 Reinforcement Learning

Reinforcement Learning, commonly referred as RL is defined in [4] as an interdisciplinary area of machine learning concerned with how an intelligent agent should choose actions in different states of a dynamic environment in order to maximize the cumulative reward. RL is one of the three basic machine learning paradigms: supervised, unsupervised and reinforcement.

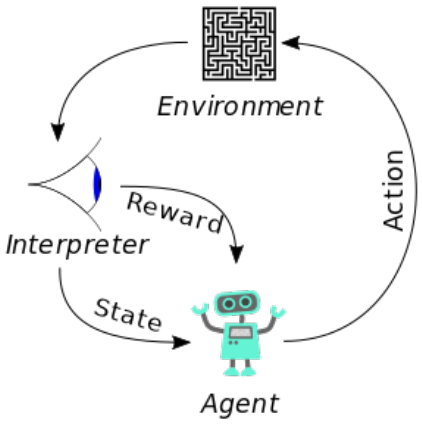

Figure 1: Typical framing of RL. [4]

The field of Reinforcement Learning is expansive, encompassing a wide range of concepts and techniques within machine learning. It is impractical to cover every aspect of RL comprehensively in this section. However, Figure 3, provided by [5], highlights the most essential concepts that serve as foundational knowledge for understanding RL effectively.

Term	Description	Pros	Cons
Model-free RL	The environment is a black box. Agents mostly conduct a trial-and-error procedure to learn on its own. They use rewards to update their decision models	The algorithm does not need a model of the environment	Requires a large amount of samples
Model-based RL	Agents construct a model that simulates the environment and use it to generate future episodes. By using the model, agents can estimate not only actions but also future states	Speed up learning and improve sample efficiency	Having an accurate and useful model is often challenging
Temporal difference learning	Use TD error to estimate the value function. For example, in Q-learning, $Q(s_t, a_t) = Q(s_t, a_t) + \beta(r_t + \gamma \max_a Q(s_{t+1}, a))$	Fast convergence as it does not need to wait until the episode ends	Estimates can be biased
Monte-Carlo method	Estimate the value function by obtaining the average of the same values in different episodes $Q(s_t, a_t) = \lim_{N->\infty} \sum_{i=1}^{N} Q(s_t^i, a_t^i)$	The values are non-biased estimates	Slow convergence and the estimates have high variances. Has to wait until episode ends to do updates
Continuous action space	The number of control actions is continuous	A policy-based method can be used	Cannot use a value-based method
Discrete action space	The number of control actions is discrete and finite	Both policy-based method and value-based method can be used	Intractable if the number of actions is large
Deterministic policy	The policy maps each state to a specific action	Reduce data sampling	Vulnerable to noise and stochastic environments
Stochastic policy	The policy maps each state to a probability distribution over actions	Better exploration	Requires a large amount of samples
On-policy method	Improve the current policy that the agent is using to make decisions	Safer to explore	May be stuck in local minimum solutions
Off-policy method	Learn the optimal policy (while samples are generated by the behavior policy)	Instability. Often used with an experience replay	Might be unsafe because the agent is free to explore
Fully observable environment	All agents can observe the complete states of the environment	Easier to solve than partially observable environments	The number of state can be large
Partially observable environment	Each agent only observes a limited observation of the environment	More practical in real-world applications	More difficult to solve as the agents require to remember past states

Figure 2: RL important concepts. [5]

Furthermore, within the realm of Reinforcement Learning, there exists a diverse array of algorithms, each offering unique approaches to problem-solving. Our aim is to explore a broad spectrum of these algorithms. A comprehensive review of various RL algorithms can be found in [5], which provides a thorough differentiation among them.

Method	Description and Advantage	Technical Requirements	Drawbacks	Implementation
Value-based method				
DQN	Use a deep convolutional network to directly process raw graphical data and approximate the action-value function	• Experience replay • Target network • Q-learning	○ Excessive memory usage ○ Learning instability ○ Only for discrete action space	[85] [86] [87] [88]
Double DQN	Mitigate the DQN's maximization bias problem by using two separate networks: one for estimating the value, one for selecting action.	• Double Q-learning	○ Inherit DQN's drawbacks	[88]
Prioritized Experience Replay	Prioritize important transitions so that they are sampled more frequently. Improve sample efficiency	•Importance sampling	○ Inherit DQN's drawbacks ○ Slower than non-prioritized experience replay (speed)	[85] [88]
Dueling Network	Separate the DQN architecture into two streams: one estimates state-value function and one estimates the advantage of each action	• Dueling network architecture • Prioritized replay	○ Inherit DQN's drawbacks	[88]
Recurrent DQN	Integrate recurrency into DQN Extend the use of DQN in partially observable environments	• Long Short Term Memory	○ Inherit DQN's drawbacks	[88]
Attention Recurrent DQN	Highlight important regions of the environment during the training process	• Attention mechanism • Soft attention • Hard attention	○ Inherit DQN's drawbacks	[68]
Rainbow	Combine different techniques in DQN variants to provide the state-of-the-art performance on Atari domain	• Double Q-learning • Prioritized replay •Dueling network •Multi-step learning •Distributional RL •Noisy Net	○ Inherit DQN's drawbacks	[85]
Policy-based method				
A3C/A2C	Use actor-critic architecture to estimate directly the agent policy. A3C enables concurrent learning by allowing multiple learners to operate at the same time	• Multi-step learning • Actor-critic model • Advantage function • Multi-threading	○ Policy updates exhibit high variance	[86] [87] [88]
UNREAL	Use A3C and multiple unsupervised reward signals to improve learning efficiency in complicated environments	• Unsupervised reward signals	○ Policy updates exhibit high variance	[89]
DDPG	Concurrently learn a deterministic policy and a Q-function in DQN's fashion	• Deterministic policy gradient	○ Support only continuous action space	[87] [88]
TRPO	Limit policy update variance by using the conjugate gradient to estimate the natural gradient policy TRPO is better than DDPG in terms of sample efficiency	• Kullback-Leibler divergence • Conjugate gradient • Natural policy gradient	○ Computationally expensive ○ Large batch of rollouts ○ Hard to implement	[87] [88]
ACKTR	Inherit the A2C method Use Kronecker-Factored approximation to reduce computational complexity of TRPO ACKTR outperforms TRPO and A2C	• Kronecker-factored approximate curvature	○ Still complex	[87]
ACER	Integrate an experience replay into A3C Introduce a light-weight version of TRPO ACER outperforms TRPO and A3C	• Importance weight truncation & bias correction • Efficient TRPO	○ Excessive memory usage ○ Still complex	[87] [88]
PPO	Simplify the implementation of TRPO by using a surrogate objective function Achieve the best performance in continuous control tasks	• Clipped objective • Adaptive KL penalty coefficient	○ Require network tuning	[86] [87] [88]

Figure 3: RL algorithms. [5]

1.1.3 Deep Learning

Deep Learning, is a class of machine learning algorithms, based on artificial neural networks (ANNs) that with the use of multiple layers, tries to progressively extract higher-level features from an input. A common example to reflect how do they work is in image processing. In lower layers maybe edges are identified while in higher layers may be that more human concepts are identified such as faces, numbers or letters. This definition is according to [6].
There are numerous types of deep learning layers, each serving a specific purpose in neural network architectures. Presented below are some of the most important types, along with their respective functions. This information is derived from [7].

- **Convolutional Layer:** Used for feature extraction in convolutional neural networks (CNNs), particularly effective for image processing tasks.
- **Dense Layer:** Found in traditional feedforward neural networks, these layers connect every neuron from the previous layer to every neuron in the current layer, allowing for comprehensive feature learning.
- **Recurrent Layer:** Integral for processing sequential data, such as time series or natural language, by preserving information from previous time steps or words.
- **Pooling Layer:** Reduces the spatial dimensions of feature maps, aiding in translation invariance and computational efficiency in CNNs.
- **Dropout Layer:** Helps prevent overfitting by randomly dropping a proportion of neurons during training, forcing the network to learn more robust representations.
- **Batch Normalization Layer:** Normalizes the activations of each layer to improve training stability and accelerate convergence.

These layers represent fundamental building blocks in deep learning models, each contributing to the network's ability to learn and generalize from data

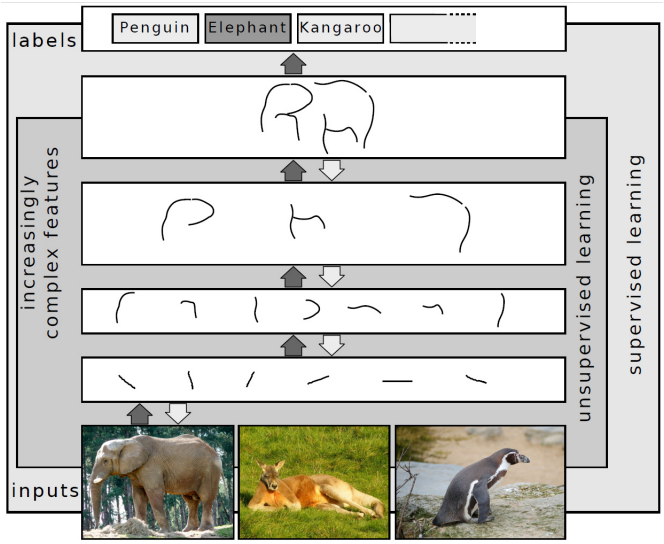

Figure 4: Image representation in multiple layers. [6]

1.2 Problem description

Addressing complexity in NP-hard problems presents a significant challenge. Many real-life problems fall into this category, rendering them intractable for large input sizes. The primary objective of this thesis is to tackle this issue by proposing novel approaches grounded in reinforcement learning (RL) techniques.

1.3 Actors involved

It's very clear to understand that the thesis is aimed for researches into the artificial intelligence field and more specifically at deep reinforcement learning researchers.They are the ones who will use the result of this project to further expand the state of the art.

The potential beneficiaries are vast, since NP-hard problems encompass a wide array of real-life scenarios. For instance, enhancing heuristics for the Traveling Salesman Problem (TSP) could yield significant benefits across numerous industries reliant on efficient routing. By enabling companies to compute more optimal routes in less time and with fewer resources, this research stands to enhance operational efficiency and resource utilization across various sectors.

1.4 State of the art

The state of the art of this field is quite extensive and a primary objective of this thesis is to comprehensively analyze these existing approaches proposed for this broad problem in order to provide with a better solution. For instance, in [8] a notable solution employing attention layers, a type of deep learning architecture, alongside a Reinforce framework, a specific reinforcement learning technique, showcases promising results across sizable instances of combinatorial optimization problems, which are inherently NP-hard.

2 Justification of the chosen alternative

Traditionally, combinatorial optimization problems have been solved using two distinct manners: with exact methods and heuristic approaches.However, a recent trend has emerged, shifting towards a novel methodology of learning heuristics through models rather than relying solely on manually crafted logic. This alternative differantiation can be found on [3].

2.1 Exact Methods

Exact methods consist on hand-written logic that solve the instance of the problem with optimality. This approach boasts the advantage of guaranteeing the best possible solution for every provided instance, ensuring optimality. However, this advantage comes at the cost of high computational complexity.

The inherent complexity of exact methods presents a notable drawback, particularly evident as the input size of the problem instances increases. As the size of the input grows, the computational demands of exact methods escalate rapidly, often rendering them infeasible for practical application.

2.2 Heuristic approach

The traditional heuristic approach similarly relies on meticulously crafted algorithms tailored to solve problem instances. However, in contrast to exact methods, heuristics prioritize computational efficiency over guaranteeing optimal solutions.

This trade-off allows heuristic methods to deliver solutions that are often close to optimal, while exhibiting significantly lower computational complexity compared to exact methods. As a result, heuristic approaches demonstrate greater scalability with respect to the size of the input instances.

2.3 Model Learned heuristics

This lastly developed approach, as previously mentioned, diverges from manually crafted logic using instead machine learning models. This methodology, precedes the encountered heuristic with a training phase enabling the model to generate nearly instantaneous solutions with a low optimality gap when well-trained.
This innovative approach further enhances computational efficiency, significantly reducing resolution times. However, this efficiency comes at the cost of diminished explainability regarding the inner workings of the heuristic.

2.4 Justification

Reinforcement Learning framework framework represents a key component of the latest alternative proposed and nd serves as the focal point of study in this thesis. Being a relatively recent development, this approach remains relatively underexplored compared to other options. However, despite its nascent status, RL has demonstrated exceptional efficacy in solving complex problems.

As such, the compelling combination of RL's promising results and its relatively unexplored terrain justifies its consideration as a viable alternative for tackling the challenges at hand.

3 Project Scope

In every project, it's important to define the scope of it given the inherent limitations of time and resources. Additionally, it's always welcome to define clearly it's objectives, requirements and possible foreseen obstacles and risks is essential to have a more robust knowledge and strategy to fulfill all the proposed goals.

3.1 Objectives

The main objective of this graduation thesis is to expand the state of the art of the of reinforcement learning applied to NP-hard problem resolution with the intention to provide new knowledge to the scientific community of artificial intelligence. Given the expansive nature of the task and the constraints of a tight timeline, a sequential objective strategy is adopted. This strategy allows for a structured approach to the research, with defined objectives at each stage, while also providing flexibility for potential endpoints along the way:

1. **Investigate, Structure, and Synthesize the Current State of the Art (O1):** This initial phase involves conducting a comprehensive review of existing literature to gain insights

into the latest developments and methodologies in reinforcement learning applied to NP-hard problems.

2. **Develop and Evaluate Different RL Frameworks for Solving Single Instances (O2):** In this phase, various reinforcement learning frameworks will be developed and evaluated for their efficacy in solving individual instances of NP-hard problems. These frameworks will be assessed based on their performance metrics and compared against each other to identify strengths and weaknesses.

3. **Select Optimal Frameworks for Generalization Capacity (O3):** Building upon the findings from the previous stage, the most promising reinforcement learning frameworks will be selected for further adaptation to handle problems with generalization capacity. This involves enhancing the frameworks to enable them to generalize solutions across a broader range of problem instances and evaluate them based on comparison against existing approaches.

4. **Provide explainability on the developed heuristics (O4):** Finally, this phase focuses on gaining a deeper understanding of how the developed heuristics operate with the ambition to provide some kind of explainability to better understand RL frameworks.

3.2 Requirements

In this section, we outline the requirements necessary to develop a new functional framework. These requirements are divided into two categories: Functional Requirements, which detail the capabilities of the framework, and Non-Functional Requirements, which specify additional characteristics and criteria beyond the functional aspects. All the concepts presented in this section are defined and extracted from [7].

3.2.1 Functional Requirements

There are some key aspects that the developed framework must fulfill to be considered functional:

- **Environment Interface (R1):** The framework should provide an interface for interacting with the environment, allowing agents to observe states, take actions, and receive rewards. This interface should facilitate seamless communication between the agent and the environment.

- **Agent Architecture (R2):** The framework should support various types of agent architectures. It should also allow for custom agent architectures tailored to specific problem domains.

- **Training Infrastructure (R3):** The framework should include training infrastructure to facilitate the training of RL agents. This infrastructure may include tools for managing training data, orchestrating training processes, and monitoring agent performance.

- **Exploration and Exploitation Strategies (R4):** The framework should support mechanisms for balancing exploration (trying out different actions to discover optimal strategies) and exploitation (leveraging known strategies to maximize rewards). It should provide options for specifying exploration strategies, such as epsilon-greedy exploration or Boltzmann exploration.

- **Parallelization (R5):** Parallelization is closely related to efficiency and plays a crucial role in optimizing training processes. By leveraging parallel computing resources, the framework can potentially achieve faster and more comprehensive training, leading to improved results.

3.2.2 Non Functional Requirements

In addition to the functional requirements outlined earlier in this section, there are also key characteristics that the developed framework must fulfill:

- **Efficiency (R6):** Efficiency is paramount for improving upon current solutions, as computational complexity is a significant challenge in solving NP-hard problems and reinforcement learning training needs as many interactions as possible to get the results. Enhancing the efficiency of the framework will enable faster and more effective problem-solving.
- **Performance (R7):** The framework should demonstrate high performance in terms of speed, efficiency, and scalability. It should be capable of handling large-scale datasets and training RL agents efficiently.
- **Robustness (R8):** The framework should be robust against noisy or incomplete input data, environmental uncertainties, and adversarial attacks. It should be able to produce reliable results under various conditions and inputs.
- **Scalability (R9):** The framework should scale efficiently with increasing problem sizes, computational resources, and training data. It should support parallelization and distributed computing to leverage multiple processors or GPUs for accelerated training.

3.3 Obstacles and Risks

Reinforcement learning frameworks, while powerful, can encounter various obstacles and risks during development and deployment. Some potential obstacles and risks include:

- **Sample Efficiency:** Reinforcement learning algorithms often require a large number of samples or interactions with the environment to learn effective policies. Limited sample efficiency can prolong training times and hinder the scalability of the framework to real-world applications.

  To address the challenge of limited sample efficiency in RL, various strategies can be employed depending on the algorithm used. Examples of approaches to improve sample efficiency include:

  - Implement experience replay to reuse past experiences during training.
  - Prioritize experiences based on their relevance.
  - Use model-based methods to leverage environment dynamics.
  - Apply transfer learning to utilize knowledge from related tasks.

- **Exploration-Exploitation Tradeoff:** Reinforcement learning frameworks must balance exploration (trying out different actions to discover optimal strategies) and exploitation (leveraging known strategies to maximize rewards). Finding the right balance between exploration

and exploitation can be non-trivial and may impact the performance of the framework.

To manage the exploration-exploitation tradeoff in RL there exist techniques that help strike a balance between exploring new actions and exploiting known strategies to maximize performance:

- Use epsilon-greedy or softmax policies for a balanced approach.
- Apply UCB methods to prioritize actions while considering uncertainty.
- Employ Thompson sampling for probabilistic action selection.
- Adapt multi-armed bandit algorithms for sequential decision-making.

- **Generalization:** Reinforcement learning frameworks may struggle to generalize well to unseen or novel environments or tasks. Generalization limitations can hinder the transferability of learned policies to new scenarios or domains.

  To address the challenge of generalization limitations in RL, several strategies can be employed to mitigate generalization limitations and improve the transferability of learned policies to new environments or tasks:

  - Ensure diverse training data.
  - Apply regularization techniques.
  - Utilize transfer learning.
  - Employ ensemble methods.
  - Introduce domain randomization.

- **Training Stability:** Reinforcement learning training processes can be unstable, especially when using deep neural network architectures. Issues such as vanishing gradients, diverging gradients, or catastrophic forgetting can disrupt training and hinder convergence to optimal solutions.

  To address the challenge of training instability in reinforcement learning, several strategies can be employed to mitigate training instability and improve the convergence of training processes towards optimal solutions:

  - Use gradient clipping to prevent exploding gradients.
  - Apply batch normalization to stabilize training.
  - Implement learning rate scheduling for smoother convergence.
  - Utilize appropriate weight initialization techniques.
  - Employ experience replay to mitigate the impact of rare events.

- **Resource Constraints:** Reinforcement learning frameworks often require significant computational resources, including high-performance computing infrastructure and large-scale datasets. Resource constraints may limit the scalability and accessibility of the framework, particularly in resource-constrained environments. In the context of RL frameworks, overcoming this challenge typically entails substantial expenses.

All the concepts presented in this section are defined and extracted from [7].

4 Methodology

This project follows a structured approach consisting of an investigation phase, followed by brainstorming sessions to generate new ideas for implementation, a development phase, and finally, an evaluation phase. This cyclical process is designed to continuously iterate and improve upon the project's outcomes over time. This methodology can be referred to as a "Cyclic Waterfall" approach, where each cycle encompasses the stages of investigation, ideation, development, and evaluation, facilitating ongoing refinement and progress.

4.1 Validation

This process will be complemented by weekly meetings, fostering open communication of project developments and advancements. During these meetings, updates and progress are going to be shared, made since the previous meeting. Additionally, each meeting will include the proposal of a set of goals to be achieved before the subsequent meeting.

During the development phases, regular validation sessions will be conducted to ensure the integrity and accuracy of the development process within the framework. These sessions will serve as checkpoints to verify that the implementation aligns with project requirements and objectives. By incorporating these validation sessions into the development workflow, the team can identify and address any issues or discrepancies early on, ultimately enhancing the quality and effectiveness of the framework.

5 Temporal Planning

This section is designed to outline a detailed timeline for the completion of the graduation thesis. The project commenced on February 15, 2024, and is scheduled for completion by June 29, 2024, culminating in the presentation and defense of the thesis.

5.1 Tasks descriptions

In this section, we will comprehensively detail the individual tasks, systematically grouped to clearly delineate the distinct phases of the project. This structured approach ensures each phase is distinctly defined, allowing for a more organized and efficient workflow. By categorizing tasks in this manner, we aim to provide clarity on the progression and dependencies of the project's various components. This organization not only facilitates a better understanding of each phase but also aids in efficient project management.

5.1.1 Study (ST)

These tasks pertain to activities that necessitate in-depth study or research on a specific topic. For this set of tasks it's only required to have internet connection and a device to search all the necessary information.

- **DRL applied to NP-HARD problems (ST1):** This task involves a comprehensive review and acquisition of knowledge about cutting-edge techniques in deep reinforcement learning that are currently employed to tackle NP-hard problems. Given the significance of this phase in the development of new heuristics, we have allocated approximately 45 hours to this task. This duration reflects the critical importance of understanding the state-of-the-art in DRL

applications for NP-hard problems, ensuring a robust foundation for the project's subsequent phases.

- **Explainability for Deep Reinforcement Learning (ST2):** This task plays a crucial role in offering insights and understanding of the heuristics developed through our project. While the primary aim is not to devise new techniques for explainability, it is still vital to ensure that the heuristics are interpretable and their workings are transparent. Consequently, we have estimated that this task will require approximately 25 hours. This time allocation is designed to adequately address the explainability aspect of our heuristics, balancing the need for clarity without diverting significant resources from the core objectives of the project.

5.1.2 Development (D)

These tasks are essential components of the thesis development phase and form a crucial segment of the project's overall activities. Given the intensive computational demands typically associated with Deep Reinforcement Learning (DRL) frameworks, these tasks are likely to require substantial computational resources to yield effective results within a reasonable timeframe. Therefore, securing the most powerful computing device available is a strategic priority, ensuring the project's computational needs are met efficiently and effectively. This approach is aimed at optimizing the performance and outcome of the research.

- **DRL frameworks for solving single instances (D1):** The development of these frameworks is crucial for evaluating and analyzing which framework is most effective for the selected problems. Recognizing the importance of this phase, we plan to invest a significant amount of time in testing various frameworks. This comprehensive approach will provide a diverse range of options for the subsequent generalization phase. Accordingly, we have allocated 45 hours to this phase, ensuring thorough exploration and assessment of different frameworks to identify the most suitable ones for our project needs. The time provided do not take into account the training phase since is very unpredictable and almost do not require human attention.

- **Selected DRL frameworks for generalization of the problem (D2):** The phase following D1 is likely to be one of the most critical stages of the project, as it involves the development of a new heuristic with generalization capabilities, which represents the one of ultimate goals of this research. Given the significance of this task, we plan to dedicate approximately 45 hours to it, ensuring enough time for the meticulous development and refinement of a heuristic that not only addresses the specific problems at hand but also demonstrates a broader applicability and effectiveness. The time provided do not take into account the training phase since is very unpredictable and almost do not require human attention.

- **Explainability framework (D3):** This task, while necessitating meticulous development, is perceived as less demanding compared to the other two development tasks, primarily because innovation is not a central component in this stage. Based on this rationale, we have decided to allocate approximately 25 hours to this task. This time frame is carefully considered to ensure sufficient attention to detail and quality, while acknowledging that the primary focus here is on execution rather than innovation.

5.1.3 Evaluation (E)

This set of tasks focuses on evaluating developments generated earlier in the project. They will likely require benchmarks and computational resources, though not as extensively as in phases D1

or D2.

- **Single instance frameworks (E1):** The evaluation of the single instance framework will involve a comparative analysis of various statistics to determine the most suitable framework for the task. As this phase is not particularly time-intensive, we anticipate completing it in under 15 hours.
- **Generalization frameworks (E2):** The evaluation of the generalization framework will encompass a comparative analysis using state-of-the-art metrics and results initially explored in ST1. We expect this task to be less time-intensive compared to others and anticipate its completion in under 15 hours.
- **Explainability framework (E3):** The evaluation of the explainability framework presents a challenge due to the less obvious nature of its metrics. However, we do not intend to allocate extensive time to this task. It is anticipated that a thorough evaluation can be efficiently completed in under 10 hours.

5.1.4 Selection (SE)

This set of tasks are the ones that are considered to be selections among various options. This tasks do not require of any extra resource.

- **Single instance (SE1):** Selecting the optimal frameworks for solving single instances, NP-Hard problems is a crucial task that heavily relies on the outcomes of D1 and E1. Provided the preceding work is executed accurately, this selection process should not require more than 5 hours.
- **Generalization (SE2):** Selecting the optimal framework, if necessary, for solving NP-Hard problems could be necessary if more than one framework is developed. This task mostly relies on the outcome of D2 and E2 and given their results is should not take more than 5 hours to select this best heuristic.

5.1.5 Analysis (A)

This group of tasks is responsible for analyzing and synthesizing ideas from various topics. Being primarily a reasoning process, it requires no resources other than the documentation generated by related tasks.

- **Heuristic (A1):** The analysis of the newly developed heuristic is a vital task, closely related to E2. It demands a deep and thorough understanding of the heuristic's intricacies. We estimate that this task will take approximately 5 hours to complete.
- **Explainability (A2):** The task of analyzing the explainability of the heuristic also presents a unique challenge, as it lacks clear metrics and requires innovative thinking. We estimate that this task will require approximately 5 hours to complete.

5.1.6 Follow up (F)

These tasks involve the essential follow-up activities for the project, ensuring thorough oversight and adherence to the schedule. The only requirement for these tasks is access to an internet connection or a suitable location for meetings.

- **Meetings (F1):** Follow-up meetings are scheduled weekly throughout the project's duration and are crucial for reviewing recent achievements and proposing new subobjectives. Each meeting is expected to last approximately one hour to have a total amount of dedicated time more or less of 25 hours.
- **Correction sessions (F2):** These sessions, particularly during the development phases, are aimed at ensuring the proper development of the respective frameworks, which can be challenging to construct. Given the emphasis on achieving the highest quality for the thesis, we estimate that these sessions will cumulatively amount to approximately 15 hours in total.

5.1.7 Documentation (DC)

These tasks are focused on documentation, requiring only a computer to compile and prepare the necessary documentation.

- **Project Management (DC1):** To maintain an organized approach, the entire project management process is meticulously documented. This ensures a clear understanding and establishment of procedures, objectives, scope, and other critical concepts vital for achieving positive outcomes and also serves as a staring point for the D2. Documenting these details is a time-intensive process, and we estimate it will take approximately 40 hours to complete. This task can be effectively broken down into several sub-tasks to provide a clearer understanding of the various components involved in the project management documentation:
  - **Contextualization & scope (DC1.1):**This sub-task is dedicated to providing a comprehensive overview of the project, including its objectives and background. The aim is to clearly define the project's purpose and its expected outcomes. This portion is anticipated to take approximately 20 hours, reflecting the depth of detail required to accurately capture the project's essence.
  - **Temporal planning (DC1.2):** This involves creating a detailed plan that outlines the scope and timeline of each task, along with the resources required for their completion. The objective is to establish a clear roadmap of the project's workflow, ensuring efficient time management and resource allocation. This part is estimated to take around 10 hours, signifying its importance in project management.
  - **Budget and sustainability analysis (DC1.3):** This sub-task focuses on assessing the financial aspects of the project, including budget allocation and cost management. Additionally, an analysis of the project's sustainability is conducted to understand its long-term viability and environmental impact. This analysis is crucial for ensuring the project's feasibility and is estimated to be completed in approximately 10 hours.
- **Memory (DC2):** This graduation thesis is grounded in an academic perspective, signifying that the project documentation holds substantial importance and impact. Acknowledging this, it is essential to emphasize that effective documentation is a meticulous and time-consuming task. It involves not only the recording of data and findings but also a comprehensive articulation of the project's objectives, methodology, theoretical framework, analysis, and conclusions.Furthermore, the documentation must be coherent, well-structured, and adhere to academic standards, ensuring it effectively communicates the research process and insights gained. Given the depth and breadth required in this documentation, we estimate that this process will require approximately 45 hours.

Group	Total Hours: 400			
ID	**Task**	**Hours**	**Resources**	**Role**
Study (ST) - Total Hours: 70				
ST1	DRL applied to NP-HARD problems	45	Internet device	Researcher
ST2	Explainability for DRL	25	Internet device	Researcher
Development (D) - Total Hours: 115				
D1	DRL frameworks for single instances	45	Computational Resources	Developer
D2	DRL frameworks for generalization	45	Computational Resources	Developer
D3	Explainability framework	25	Computational Resources	Developer
Evaluation (E) - Total Hours: 40				
E1	Evaluation of single instance frameworks	15	Benchmarks, Computational Resources	Evaluator
E2	Evaluation of generalization frameworks	15	Benchmarks, Computational Resources	Evaluator
E3	Evaluation of explainability framework	5	Benchmarks	Evaluator
Selection (SE) - Total Hours: 10				
SE1	Selection of frameworks for single instances	5	-	Decision Maker
SE2	Selection of generalization frameworks	5	-	Decision Maker
Analysis (A) - Total Hours: 10				
A1	Analysis of the heuristic	5	Documentation	Analyst
A2	Analysis of explainability	5	Documentation	Analyst
Follow up (F) - Total Hours: 40				
F1	Follow-up meetings	25	Internet, Meeting Space	Project Manager
F2	Correction sessions	15	Internet, Meeting Space	Quality Assurance
Documentation (DC) - Total Hours: 85				
DC1.1	Contextualization & Scope	20	Computer	Documenter
DC1.2	Temporal Planning	10	Computer	Documenter
DC1.3	Budget and Sustainability Analysis	10	Computer	Financial Analyst
DC2	Project Documentation	45	Computer	Documenter

Table 1: Project Tasks Overview. Own elaboration.

5.2 PERT & Gantt charts

To offer a clearer perspective on the workflow and timing of the thesis, we present a PERT chart for a comprehensive visualization of task dependencies and a Gantt chart to track task progression over time. The PERT chart exposes the interdependencies of tasks, providing a roadmap for project execution, while the Gantt chart delineates the temporal development of tasks, articulated on a weekly basis. These tools are instrumental in streamlining project management and are referenced in Figures 5 and 6, respectively.

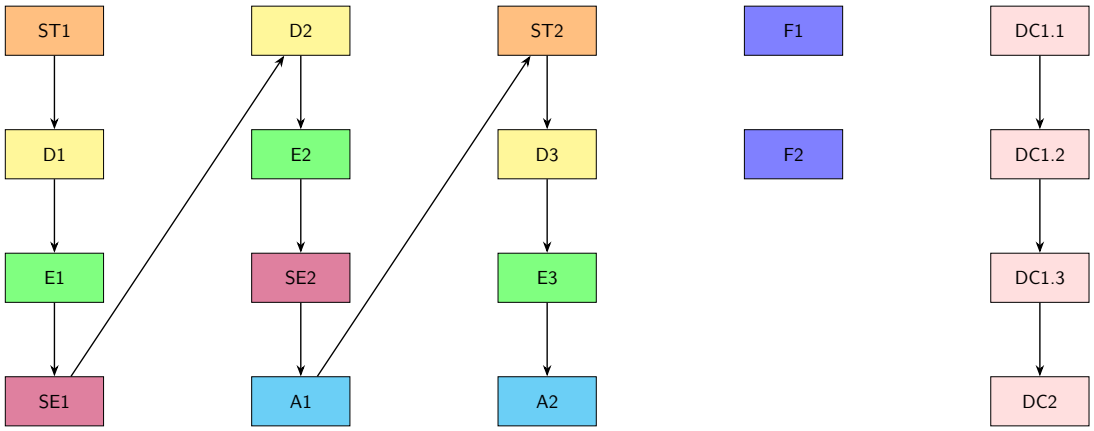

Figure 5: Task Dependency Chart. Own elaboration.

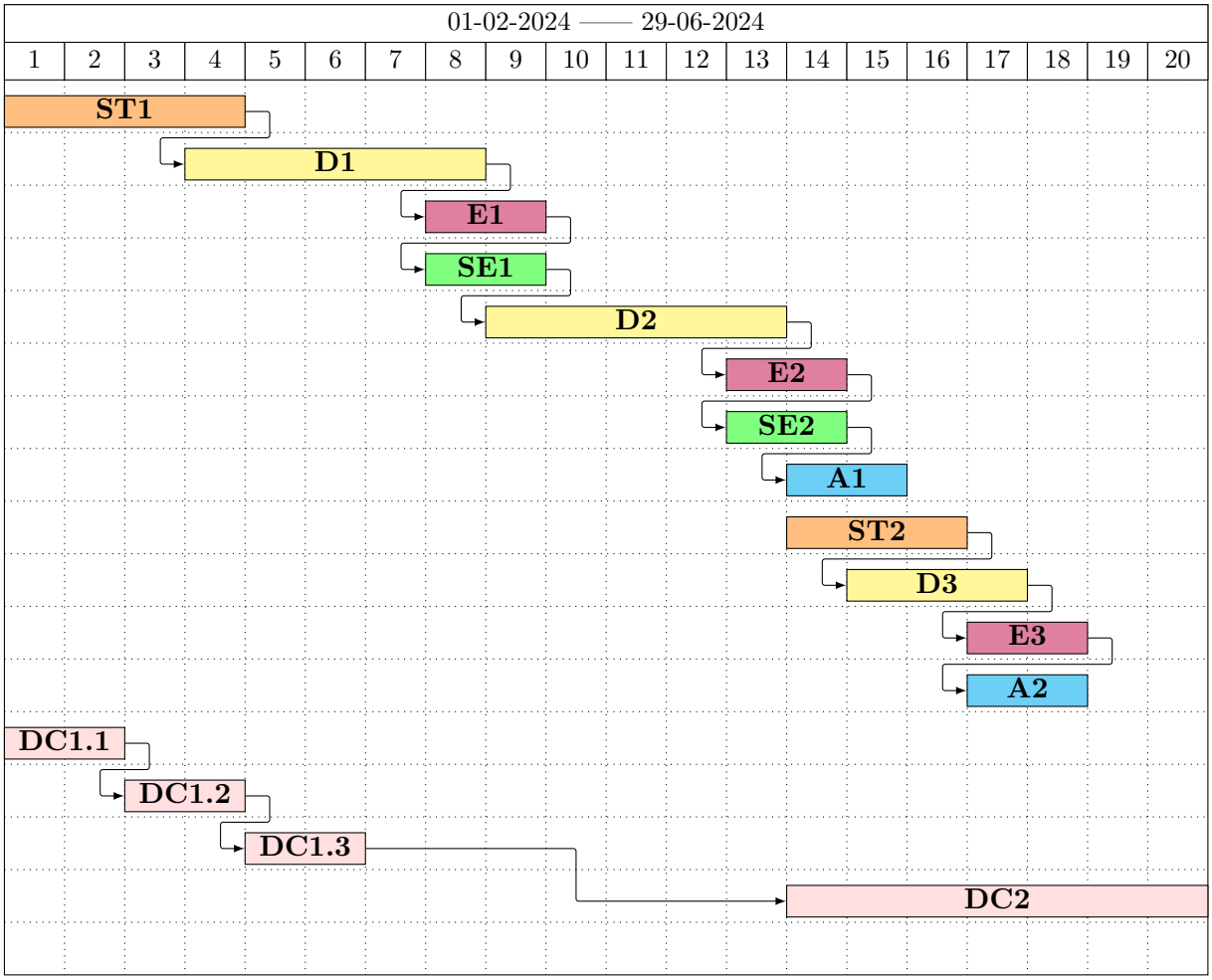

Figure 6: Gantt diagram. Own elaboration.

5.3 Risk management: Challenges and Mitigation Strategies in Reinforcement Learning Frameworks

Reinforcement learning (RL) frameworks are powerful tools but can face several challenges and risks during development and deployment. Since this is the main source of problems that we can face in this thesis, here is provided a list of possible obstacles and how will they be tackled:

- **Sample Efficiency:** RL algorithms often require numerous samples to learn effective policies. This inefficiency can extend training durations and limit real-world scalability.

  Mitigation Strategies:

  - Utilize experience replay to recycle past interactions.
  - Prioritize experiences by significance.
  - Employ model-based methods to harness environment dynamics.
  - Adopt transfer learning from related tasks.

- **Exploration-Exploitation Tradeoff:** RL must balance the act of exploring new actions with exploiting known strategies to maximize rewards, which is a nuanced challenge.

Mitigation Techniques:

  - Implement epsilon-greedy or softmax policies.
  - Use Upper Confidence Bound (UCB) methods for action prioritization.
  - Engage in Thompson sampling for action selection.
  - Adapt multi-armed bandit algorithms for decisions.

- **Generalization:** RL frameworks may not generalize effectively to new environments or tasks, affecting policy transferability.

  Strategies to Improve Generalization:

  - Ensure training on diverse datasets.
  - Apply regularization techniques.
  - Use transfer learning for broader applicability.
  - Employ ensemble methods and domain randomization.

- **Training Stability:** The RL training process can be volatile, particularly with deep neural networks, leading to issues like vanishing or exploding gradients.

  Mitigation Strategies:

  - Utilize gradient clipping to prevent gradient issues.
  - Apply batch normalization for stable training.
  - Implement learning rate scheduling for better convergence.
  - Use suitable weight initialization and experience replay.

- **Resource Constraints:** RL frameworks demand substantial computational power and datasets, which can be prohibitive in resource-limited settings.

  Overcoming Resource Constraints:

  - The solutions here often involve significant investment in computational infrastructure.

The challenges associated with Reinforcement Learning (RL) frameworks have been considered and are reflected in the time allocated to each task within the project. Concerning resource constraints, we are proactively exploring solutions to acquire powerful computational resources at minimal cost. Our strategy includes seeking partnerships with entities equipped with high-performance computing facilities. One potential collaborator could be the Barcelona Supercomputing Center, which could offer access to advanced computational capabilities. By leveraging such partnerships, we aim to mitigate the limitations imposed by resource constraints, thereby ensuring the scalability and success of our RL frameworks without incurring excessive expenses. This forward-thinking approach is designed to align with our commitment to efficiency and fiscal responsibility throughout the development of the project.

6 Economic management and sustainability

This section is structured to meticulously outline the economic management and sustainability aspects of the graduation thesis. It commences with a comprehensive overview of the project's budget, including a detailed identification and estimation of all associated costs. This is followed by a thorough analysis of management control mechanisms, ensuring efficient resource allocation and financial oversight. The document culminates with an in-depth sustainability report, evaluating the project's long-term viability and its alignment with broader economic, social, and environmental objectives.

6.1 Economic management

To ensure effective economic management, it is crucial to comprehensively define all aspects related to the budget of this project. This takes into account, as previously stated, the identifiaction and estimation of the costs and management control of it.

6.1.1 Identification and estimation of costs

Building on the budget planning framework outlined in [9], our costs can be categorized into four distinct groups. This classification simplifies the process and proves beneficial for identifying all expenses. By structuring the budget planning into these separate categories, we gain a clearer and more organized overview of our financial landscape.

- **Personnel costs:** These costs are associated with the personnel involved in the project. Since each role carries a distinct market value, it's imperative to identify all the roles within the project and assign competitive salaries accordingly. Moreover, we must account for social security (SS) contributions, which effectively means multiplying the gross salary by 1.3 to cover these additional expenses. In Table 2 we have an overview for all different roles of the gross salary, the social security payment and the overall retribution. It's important to remark that all the salaries are based on average salaries in Spain by [10, 11].

Role	**Gross Salary**	**SS**	**Retribution**
AI Researcher	20€/h	6€/h	26€/h
AI Developer	30€/h	9€/h	39€/h
Decision Maker	50€/h	15/h	65€/h
Analyst	20€/h	6€/h	26€/h
Project Manager	32.5€/h	9.75€/h	42.25€/h
Quality Assurance	21€/h	6.33€/h	27.33€/h
Documenter	16€/h	4.8€/h	20.8€/h
Financial Analyst	30€/h	9€/h	39€/h
Evaluator	25€/h	7.5€/h	32.5€/h

Table 2: Cost breakdown per role. Own elaboration.

Upon compiling a comprehensive overview of the total compensation for each role involved in the thesis, we can then proceed to calculate the CPA (cost per activity). This is accomplished by estimating the costs for each task outlined in the Gantt chart. Table 3 provides a detailed breakdown of these tasks, including their estimated costs, thus offering a deep understanding of the financial implications associated with each activity.

ID	Task	Hours	Resources	Role	Cost (€)
Study (ST) - Total Hours: 70					
ST1	DRL applied to NP-HARD problems	45	Internet device	Researcher	1170
ST2	Explainability for DRL	25	Internet device	Researcher	650
Development (D) - Total Hours: 115					
D1	DRL frameworks for single instances	45	Computational Resources	Developer	1755
D2	DRL frameworks for generalization	45	Computational Resources	Developer	1755
D3	Explainability framework	25	Computational Resources	Developer	975
Evaluation (E) - Total Hours: 40					
E1	Evaluation of single instance frameworks	15	Benchmarks, Computational Resources	Evaluator	487.5
E2	Evaluation of generalization frameworks	15	Benchmarks, Computational Resources	Evaluator	487.5
E3	Evaluation of explainability framework	10	Benchmarks	Evaluator	325
Selection (SE) - Total Hours: 10					
SE1	Selection of frameworks for single instances	5	-	Decision Maker	325
SE2	Selection of generalization frameworks	5	-	Decision Maker	325
Analysis (A) - Total Hours: 10					
A1	Analysis of the heuristic	5	Documentation	Analyst	130
A2	Analysis of explainability	5	Documentation	Analyst	130
Follow up (F) - Total Hours: 40					
F1	Follow-up meetings	25	Internet, Meeting Space	Project Manager	1062.5
F2	Correction sessions	15	Internet, Meeting Space	Quality Assurance	409.95
Documentation (DC) - Total Hours: 85					
DC1.1	Contextualization & Scope	20	Computer	Documenter	416
DC1.2	Temporal Planning	10	Computer	Documenter	208
DC1.3	Budget and Sustainability Analysis	10	Computer	Financial Analyst	390
DC2	Project Documentation	45	Computer	Documenter	936
				Total Hours:	**400**
				Total CPA:	**11937.45**

Table 3: Detailed Cost Estimation per Task. Own elaboration.

We also provide in Table 4 a detailed overview of the budget per each group of tasks to have a more comprehensive visualization.

ID	Task	Hours	Role	Cost (€)
ST	Study	70	Researcher	1820
D	Development	115	Developer	4485
E	Evaluation	40	Evaluator	1300
SE	Selection	10	Decision Maker	650
A	Analysis	10	Analyst	260
F	Follow up	40	Project Manager, Quality Assurance	1472.45
DC	Documentation	85	Documenter, Financial Analyst	1950
			Total CPA:	**11937.45**

Table 4: Simplified Cost Estimation per Task Category. Own elaboration.

- **Generic Costs:** Generic costs refer to the necessary expenditures associated with the project that are not directly assignable to specific tasks. These encompass in this project overhead expenses such as reimbursements, workspace rental, electricity usage, and internet service charges.
  - **Reimbursements:** There is only hardware reimbursements since all the software used is open source an do not require any inversion. For the hardware, we account for a laptop valued at 1100€ and a monitor at 150€. Based on a standard operational period of 220 workdays annually, with 8 working hours per day, and factoring in the hardware's estimated lifespan of 4 years, we determine the reimbursement cost using the following formula:

$$Reimbursements = \frac{Cost * hours\ used}{Useful\ life * 220 * 8}$$

Given that the hardware is estimated to be in use for 1000 hours, primarily due to the extensive time required for model training, the total reimbursement cost is calculated to be **177.55€**.

- **Work space rental:** The project's primary workspace will be located in Gracia, a district in Barcelona. The average rent there is approximately 1100€ per month, as stated in [12]. We also need to take into account that the size of this rental is $80m^2$ with 3 rooms and 2 baths. As we only need 1 room and 1 bathroom we estimate that is more or less the 40% of the original rental price. Given the project's duration of about five and a half months, the total rental expenditure is expected to surpass **2420€**.
- **Electricity usage:** Given that nowadays the electricity price in Spain is about 0.2966€/kWh and the average power consumption of a laptop and screen are about 0.075kWh and 0.0225kWh respectively, and the previously estimated hardware usage is about 1000 hours, the total electrical expenditures extend to **29.92€**. All this prices are found on [13, 14, 15].
- **Internet service charges:** We also have to take into account the internet service charges, which are on average 45€ per month according to [16]. Since the project has a duration of 5 and a half months as previously mentioned, the total internet expenses will round up upon **247.5€**.

In Table 5, a comprehensive overview of generic costs is presented, offering a summarized visualization of the various expenses associated with the project.

Expense Category	**Amount (€)**
Reimbursements	177.55
Workspace Rental (Gracia, Barcelona)	2420.00
Electricity Usage	29.92
Internet Service Charges	247.5
Total Generic Costs	**2874.97**

Table 5: Summary of Generic Expenses. Own elaboration.

Outlining the potential collaboration with the Barcelona Supercomputing Center (BSC) is crucial, as it would significantly alter the expenses associated with electricity usage and hardware reimbursements. With the bulk of computational demands transferred to BSC's resources, the reliance on our own laptop for intensive tasks would be substantially reduced. This shift necessitates a readjustment of costs, factoring in the power usage specific to the BSC's infrastructure rather than our personal hardware.

- **Contingencies:** In any project, particularly one involving AI, it's crucial to anticipate potential cost overruns due to unforeseen challenges. For this project, a standard approach is to allocate an additional 15% to the total costs, encompassing both the CPA (cost per activity) and generic costs (GC). This contingency measure helps ensure financial preparedness for any unexpected issues that may arise during the project's lifecycle.
- **Unexpected costs:** All foreseeable challenges and risks associated with this thesis, such as those commonly encountered in deep reinforcement learning frameworks, including sample efficiency, exploration-exploitation tradeoff, and generalization issues upon others, are

already factored into the project's timeline with corresponding time extensions. Considering that these potential delays and their high likelihood have been proactively included in the planning stage, we anticipate no unexpected costs arising from these factors. This proactive approach ensures that our time planning comprehensively addresses the typical obstacles inherent in such projects.

There are only unexpected costs related with the breaking of the computer and the screen. Since this events a rarely likely to happen. We can associate a 5% probability to this events. This would result into:

$$cost * probability = (1100 + 150) * 0.05 = 62.5€$$

At Table 6 we conclude with an overview of the total costs of the project separated in each important cost category, which ensures a general overview upon the financial aspects of the thesis.

Cost Category	**Amount (€)**
Total CPA (Cost per Activity)	11937.45
Total Generic Costs (GC)	2874.97
Contingency (15%)	2221.86
Unexpected Costs	62.5
Total Costs	**17096.78**

Table 6: Summary of Total Project Costs. Own elaboration.

6.1.2 Management Control

In our project, we propose and thoroughly describe excellent mechanisms to control any budget deviations that might arise. This includes defining numerical indicators for calculating deviations, which will facilitate effective monitoring and control. Specifically, we will implement a variance analysis system, comparing actual expenditures against the planned budget periodically. This will enable us to promptly identify and analyze any discrepancies.

To quantify these deviations, we'll calculate both the variance in absolute terms and as a percentage of the budgeted amount. These indicators will be critical for understanding the scale and impact of any deviation. Regular reviews will be scheduled to ensure timely responses to these variances.

Furthermore, we plan to use a rolling forecast model. This approach allows us to adjust our forecasts based on actual performance periodically, thus improving the accuracy of our budget predictions and enhancing our ability to manage resources effectively.

In summary, our approach to budget control combines proactive planning, regular monitoring, and adaptive forecasting to ensure that any deviations are quickly identified, analyzed, and addressed, keeping the project on track financially.

6.2 Sustainability report

Sustainability is a crucial factor in all project undertakings due to its significant impact on the economy, society, and the environment. This importance necessitates a comprehensive report for this

project, derived from a carefully conducted survey, to ensure heightened awareness and integration of sustainability practices within this graduation thesis.

6.2.1 Introspection

After having answered to the survey leaded into the academic world from EDINSOST, I've gained clearer insight into my competencies and deficiencies in this area. My understanding of sustainability is fundamental, and I recognize that my experience in developing various techniques or metrics in this field is limited. Despite this, I feel well-informed about our society's sustainability challenges. However, my practical experience is insufficient to fully back this claim.

Regarding my research-focused project, I acknowledge that there is limited scope for applying sustainability measures. Nonetheless, I bear the responsibility to monitor and endeavor to reduce my environmental impact throughout the execution of my thesis. There is also really important to bear in mind the impact of the thesis within the scope of sustainability, which if the research comes to fruition, could improve the condition over the economical, social and environmental dimensions.

6.2.2 Economic dimension

Focusing on the economic dimension, it's crucial to deliberate on the estimated cost of the project. Given that the tasks are meticulously defined—which is challenging in a research project—and the costs are thoroughly justified, it seems that the financial aspects are well outlined and adhere to standard procedures.

There are a lot of ways to solve NP-HARD problems but we are focusing on deep reinforcement learning frameworks which is an heuristic approach. Historically has been solved in many different ways and is a field that evolved quite fast. In the actuality, one of the best ways to solve it is using the [8] which tries to solve the problem using a REINFORCE train framework over attention networks using a similar structure to the transformers proposed in the famous paper [17].

Moreover, the economic impact of this graduation thesis warrants attention. Although it may not have an immediate direct effect, the potential for a significant indirect impact is noteworthy. For instance, if this thesis succeeds in enhancing the solutions for TSP variants, which are emblematic NP-HARD problems, it could encourage enterprises to adopt our methodologies. This adoption could yield improved routing solutions, achieving efficiency in addressing larger problems more swiftly. However, there is a caveat: the necessity for these companies to modify their existing systems incurs an initial cost, despite the promise of long-term benefits.

6.2.3 Environemental dimension

In this section, we delve into the environmental dimension of the sustainability report, focusing on the ecological impact associated with conducting this graduation thesis. We commit to rigorously monitoring and measuring our carbon footprint throughout the thesis, particularly during the training phases, and also in everyday tasks. Our primary source of emissions is the use of our laptop, which will be meticulously monitored to accurately gauge its emissions. Furthermore, should our collaboration with the Barcelona Supercomputer Center (BSC) materialize, we pledge to similarly track the emissions generated by their infrastructure.

Efforts to mitigate our environmental impact include implementing energy-saving practices such as shutting down the computer and internet when not in use, optimizing the computer's power management, and performing regular maintenance and component cleaning to extend its useful life.

As previously mentioned, there are already effective methods for medium-sized graph problems. Our goal, however, is to enhance the quality and scalability of these solutions. This endeavor indirectly benefits numerous routing challenges, many of which are NP-HARD, potentially leading to a significant reduction in the carbon footprint of entire companies on a larger scale. This project, therefore, not only contributes to academic knowledge but also offers tangible environmental benefits.

6.2.4 Social dimension

In this section, we address the social and personal impacts of the graduation thesis. The personal impact is especially significant. Having chosen this topic out of deep interest, my dedication to both the process and the outcomes of this thesis is immense. My longstanding fascination with AI, combined with a more recent intense focus on reinforcement learning, means this project is not just an academic pursuit, but a chance to deepen my own understanding and passion in a field that excites me.

The social impact of this thesis is also substantial. As a research project, achieving our set goals will contribute to advancing the state of the art, which is always beneficial for both science and society. This progress is particularly meaningful in the context of reinforcement learning, a field that is increasingly critical across various global industries. By pushing the boundaries of current knowledge in this area, this thesis has the potential to offer new insights and solutions, especially the routing industry or biology community, where there is a growing demand for innovation and improvement. This contribution to the scientific community not only elevates the field but also supports practical advancements in industries that rely on these technologies.3

The sustainability matrix presented in 7 summarizes the key aspects of the graduation thesis in terms of economic, environmental, and social dimensions.

Dimension	Impact	Strategy	Outcome
Economic	Potential for significant indirect economic impacts through the enhancement of TSP variant solutions.	Focus on deep reinforcement learning frameworks for solving NP-HARD problems, with cost-effective and efficient methodologies.	Encouragement of enterprise adoption, leading to more efficient problem-solving and long-term financial benefits.
Environmental	Carbon emissions from computer use and potential supercomputer collaborations.	Implement energy-saving practices, monitor emissions, and optimize computational resources.	Reduction in carbon footprint for both the individual project and, potentially, on a larger scale through improved routing challenges solutions.
Social	Advancement in the field of AI and reinforcement learning, with personal growth and broader scientific contributions.	Dedication to exploring and expanding knowledge in AI, specifically reinforcement learning, to tackle significant challenges.	Contribution to scientific progress and practical advancements in industries reliant on AI and routing solutions, benefiting society at large.

Table 7: Sustainability Matrix. Own elaboration.

References

[1] Wikipedia. *NP-hardness*. https://en.wikipedia.org/wiki/NP-hardness. Accessed on February 21, 2024.

[2] Yunhao Yang and Andrew Whinston. "A survey on reinforcement learning for combinatorial optimization". In: *2023 IEEE World Conference on Applied Intelligence and Computing (AIC)*. IEEE. 2023, pp. 131–136.

[3] Leo Ardon. "Reinforcement Learning to Solve NP-hard Problems: an Application to the CVRP". In: *arXiv preprint arXiv:2201.05393* (2022).

[4] Wikipedia. *Reinforcement Learning*. https://en.wikipedia.org/wiki/Reinforcement_learning. Accessed on February 21, 2024.

[5] Ngoc Duy Nguyen et al. "Review, analysis and design of a comprehensive deep reinforcement learning framework". In: *arXiv preprint arXiv:2002.11883* (2020).

[6] Wikipedia. *Deep Learning*. https://en.wikipedia.org/wiki/Deep_learning. Accessed on February 21, 2024.

[7] Dr. J.W. Böhmer and Dr. M.T.J. Spaan. *Deep Reinforcement Learning (CS4400) Course Notes*. University of TU Delft. Unpublished course notes. 2024.

[8] Wouter Kool, Herke Van Hoof, and Max Welling. "Attention, learn to solve routing problems!" In: *arXiv preprint arXiv:1803.08475* (2018).

[9] Departament d'Organització d'Empreses. *Mòdul 2.4: Gestió Econòmica*. Facultat d'Informàtica de Barcelon (FIB). Unpublished course notes. 2024.

[10] PayScale. *Salary Data and Compensation Analysis*. https://www.payscale.com. Accessed on March 9, 2024. 2024.

[11] SalaryExpert. *Salary Data and Compensation Analysis*. http://www.salaryexpert.com/. Accessed on March 9, 2024. 2024.

[12] La Vanguardia. *Los propietarios de vivienda en Gracia, los más afortunados de Barcelona*. https://www.lavanguardia.com/economia/finanzas-personales/20230726/9133205/propietarios-vivienda-gracia-mas-afortunados-barcelona-mkt-hfy.html. Accessed on March 10, 2024. 2024.

[13] Electricity In Spain. *Electricity Prices In Spain*. https://electricityinspain.com/electricity-prices-in-spain/. Accessed on March 10, 2024. 2024.

[14] Energuide. *How much power does a computer use and how much CO2 does that represent?* https://www.energuide.be/en. Accessed on March 10, 2024. 2024.

[15] ITpedia. *Our Computer, Monitor, and Energy Consumption*. https://en.itpedia.nl/. Accessed on March 10, 2024. 2023.

[16] Remote Year. *Cost of Living in Spain*. https://www.remoteyear.com/blog/cost-of-living-in-spain. Accessed on March 10, 2024. 2024.

[17] Ashish Vaswani et al. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).