Monitoring report

Tomàs Osarte Segura

May 2024

Contents

List of Figures

List of Tables

Acronyms

AI Artificial Intelligence. 6

COP Combinatorial optimization problems. 6

DRL Deep Reinforcement Learning. 6, 16

FIB Facultat d'Informàtica de Barcelona. 6

ML Machine Learning. 6

NPH NP-Hard. 6

RL Reinforcement Learning. 6, 12–15

1 Contextualization of the project

This type A graduation thesis is situated in the scope of the Facultat d'Informàtica de Barcelona (FIB) It belongs to the computing mention within the degree and more specifically is working on an specific field of Artificial Intelligence (AI) which is Reinforcement Learning (RL).

This thesis aims to extend and improve works such as [1, 2, 3]. This papers provide different Deep Reinforcement Learning (DRL) frameworks that are used to solve different Combinatorial optimization problems (COP) that are NP-Hard (NPH).

Machine Learning (ML) methods offer several advantages into solving NPH problems, such as enhanced efficiency, real-time decision-making capabilities, customized solutions, and scalability. However, as highlighted in [4], there are several promising research directions aimed at unlocking the full potential of ML methods for solving complex problems. These include:

- **Generalization:** Enhancing the adaptability of ML frameworks to perform consistently across varied scenarios.
- **Improving Models and Solution Methods:** Developing more sophisticated models and refining techniques to enhance performance.
- **Incorporating Uncertainty and Online Routing:** Integrating strategies for managing uncertainty and enabling real-time decision-making within ML frameworks.

This study tackles the first two research directions by investigating various Deep Reinforcement Learning (DRL) frameworks. We empirically determine which frameworks are most effective for NP-Hard (NPH) problems. Additionally, this research aims to develop a new DRL framework that not only exhibits enhanced generalization capabilities but also significantly boosts overall performance across diverse scenarios.

2 Planification

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.

2.1 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.

2.1.1 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.

2.1.2 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.

2.1.3 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.

2.1.4 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.

2.1.5 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.

2.1.6 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.

2.1.7 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	70	Internet device	Researcher
Development (D) - Total Hours: 120				
D1	DRL frameworks for single instances	60	Computational Resources	Developer
D2	DRL frameworks for generalization	60	Computational Resources	Developer
Evaluation (E) - Total Hours: 40				
E1	Evaluation of single instance frameworks	20	Benchmarks, Computational Resources	Evaluator
E2	Evaluation of generalization frameworks	20	Benchmarks, Computational Resources	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	10	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.

2.2 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 1 and 2, respectively.

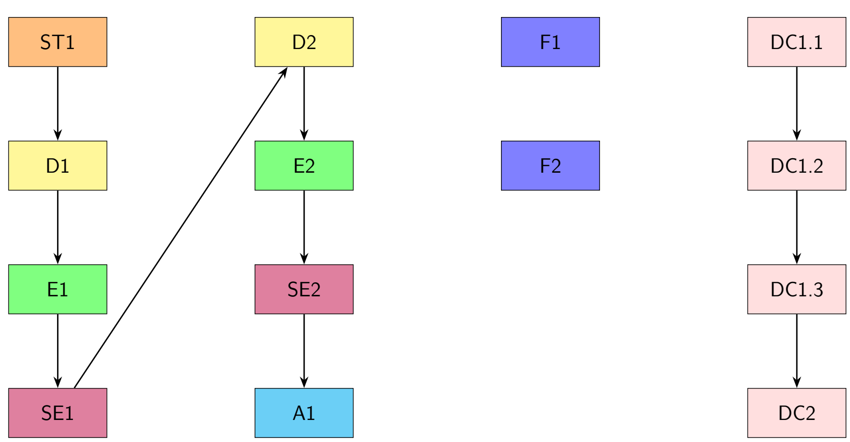

Figure 1: Task Dependency Chart. Own elaboration.

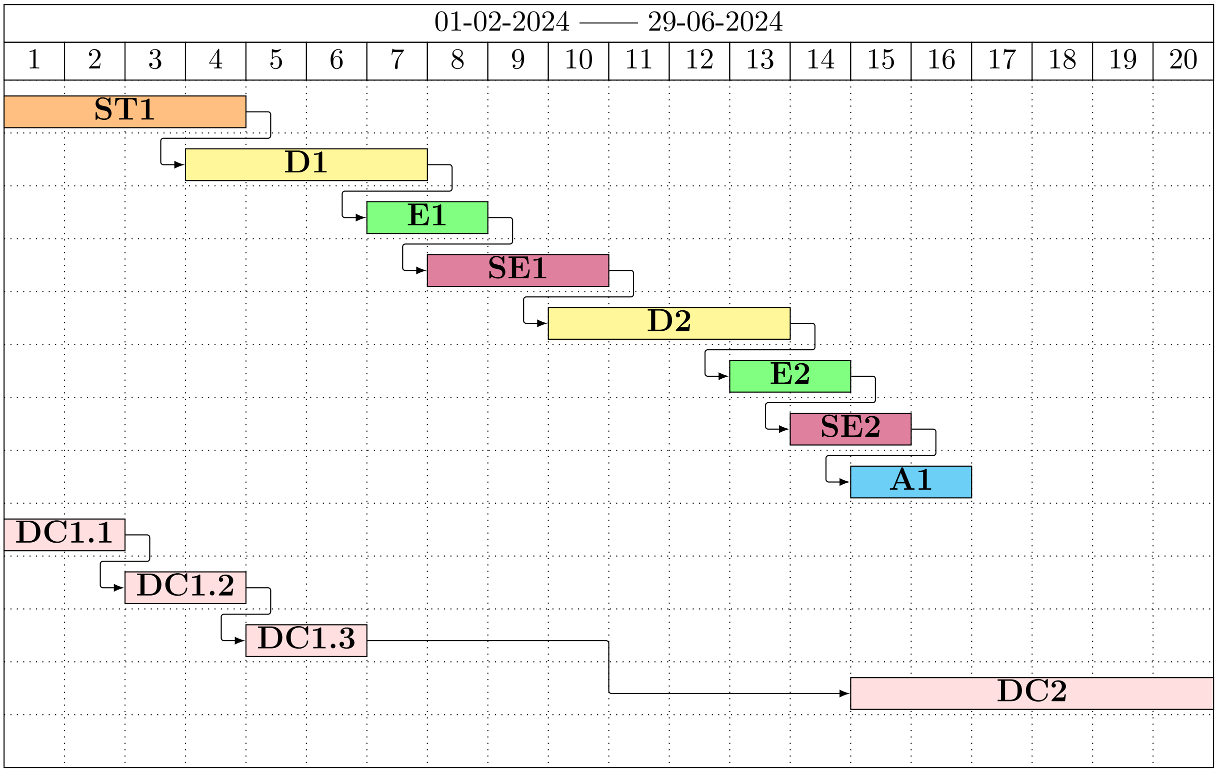

Figure 2: Gantt chart. Own elaboration.

2.3 Risk management: Challenges and Mitigation Strategies in RL Frameworks

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 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.

3 Methodology and rigor

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.

3.1 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.

4 Analysis of alternatives

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 [5].

4.1 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.

4.2 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.

4.3 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.

4.4 Justification

Reinforcement Learning framework framework represents a key component of the latest alternative proposed and 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.

5 Current thesis

From the beginning until now, in mid-May 2024, significant progress has been made in developing the thesis. This section will briefly explain the current status of the project, considering the initial planning:

- **ST1**: This stage has been completed with the study of various papers. While not all papers are mentioned here, some key ones are highlighted. The state of the art for solving combinatorial optimization problems and NP-hard problems using Deep Reinforcement Learning (DRL) involves using a reinforcement framework with a baseline and a network utilizing attention mechanisms, as presented in [6]. Subsequent advancements have incorporated specialized techniques tailored to different types of problems, as seen in [2], as well as more general sophisticated techniques found in [2, 7].

- **D1**: This phase has also been concluded with the development of a framework capable of employing various DRL learners and an infinite number of environments step by step, primarily based on [8]. Specifically, the environment and network for solving the Traveling Salesman Problem (TSP) have been developed.

- **E1 and SE1**: In this evaluation phase, different learners such as REINFORCE, REINFORCE with baseline, Actor-Critic, PPO learner, and others were tested. Based on the results, we concluded that the most effective learner for solving these types of problems is the PPO learner. However, this evaluation includes more intrinsic details that will be elaborated upon in the memory.

Currently, we are finalizing the D2 phase, which is promising in terms of potential generalization by training with diverse sizes of the TSP problem, a major challenge in the current state of the art. Future steps include testing the model in a supercomputer center to evaluate its learning capacity and benchmarking it against various state-of-the-art heuristics.

6 Knowledge Integration

During this graduation thesis, various types of knowledge acquired throughout the degree have been integrated. The knowledge learned throughout the career and used to develop this thesis can be divided into two main categories: technical and procedural.

6.1 Technical Knowledge

Throughout the degree, we have been taught numerous technical concepts that have been crucial in developing the thesis. Although it is challenging to mention all of them here, as many concepts have been relevant to this work, the most significant ones are:

- **Software Architecture:** Understanding software architecture is essential for providing a framework capable of testing different Deep Reinforcement Learning (DRL) frameworks and ensuring scalability to various instances of NP-HARD problems. A thorough grasp of software architecture concepts has been crucial.

- **Artificial Intelligence Techniques:** Knowledge of AI techniques, particularly Reinforcement Learning (RL) and Deep Learning (DL), has been vital in developing different DRL frameworks. These concepts enabled the construction and improvement of frameworks and inspired new ideas for generalization.

- **Parallelization:** The concept of parallelization has been significant in enhancing the efficiency of training processes. Understanding how to effectively parallelize tasks has contributed to optimizing the overall workflow.

6.2 Procedural Knowledge

Beyond technical aspects, the degree program has imparted essential soft skills necessary for the successful completion of this thesis. These include:

- **Project Management:** Learning how to plan, execute, and manage a project has been critical. This includes setting milestones, managing time efficiently, and coordinating various tasks and resources effectively.

- **Research Methodology:** Understanding how to conduct research, including literature reviews, hypothesis formulation, experimentation, and data analysis, has been fundamental in developing a rigorous and well-supported thesis.

- **Problem-Solving Skills:** The ability to approach complex problems systematically and develop innovative solutions has been cultivated throughout the degree program. This skill has been particularly valuable in addressing challenges encountered during the thesis development.

- **Communication Skills:** Effective communication, both written and oral, has been emphasized. These skills are essential for presenting research findings clearly and concisely, as well as for collaborating with peers and mentors.

- **Teamwork and Collaboration:** Working collaboratively with others, sharing ideas, and contributing to a team environment have been integral parts of the learning experience. These skills have facilitated productive interactions and enhanced the quality of the thesis.

The integration of these technical and procedural knowledge areas has been fundamental in the successful development of this thesis, demonstrating the comprehensive and multidisciplinary nature of the degree program.

7 Identification of Laws and Regulations

The recent ratification of the first comprehensive law on Artificial Intelligence (AI) by the European Union introduces significant regulatory measures for the use and development of AI technologies. This law, while groundbreaking, primarily targets systems that have broad public and commercial applications, particularly those involving personal data, critical infrastructure, or applications that pose a high risk to safety and fundamental rights [9, 10].

In the context of this graduation thesis, which focuses on developing and improving Deep Reinforcement Learning (DRL) frameworks for solving NP-Hard problems, the direct implications of this law are minimal. The primary reasons are as follows:

- **Scope of Application:** The EU AI law targets AI systems that are deployed in public domains or used in ways that significantly impact individuals' rights and freedoms. Since this thesis is an academic research project with no immediate plans for public deployment or commercial use, it falls outside the primary scope of the law.

- **Research Exemption:** The law includes provisions that exempt research and development activities from some of its more stringent requirements, recognizing the importance of fostering innovation. Academic research, such as the work conducted in this thesis, is generally covered under these exemptions [11].

Therefore, while the new AI law represents a significant regulatory framework, it does not impose additional constraints on this specific research project. The focus remains on advancing the state-of-the-art in DRL methodologies within the bounds of academic research and ethical practice. This ensures that the project aligns with both current legal standards and the broader goals of responsible AI development.

8 Involvement and Decision-Making

Throughout this project, I have been actively and consistently involved, particularly in my interactions with my supervisor. It has been a pleasure to work with him on this interesting project. I attended all meetings punctually, most of which I scheduled on my own initiative. Regular communication and feedback were prioritized to ensure the project's successful progression.

My enthusiasm for the project is reflected in my proactive approach and respectful interactions. I remained highly motivated and dedicated, consistently working towards and often exceeding the project's initial objectives. This commitment and thorough preparation at each milestone have ensured that the project stayed on track and achieved the desired outcomes.

I employed rigorous scientific and technical procedures to ensure the reliability and validity of the research findings. By adhering to established methods, I produced credible and high-quality results.

In summary, my proactive involvement, motivation, respectfulness, and commitment to rigorous scientific methods have been key to the successful development and completion of this project.

9 Initiative and decision-making

Throughout this project, I have maintained frequent contact with my supervisor, often discussing various aspects of the work. I have consistently brought forth interesting proposals and maintained a satisfactory work pace. I have a clear understanding of the alternative solutions available and have carefully evaluated them. I can thoroughly justify why the chosen solutions were selected.

Regular communication with my supervisor has been a cornerstone of this project's success. I actively engaged in discussions, sharing my ideas and receiving valuable feedback. This collaborative approach allowed me to refine my proposals and ensure they were well-aligned with the project's goals.

I took a methodical approach to evaluate alternative solutions. By considering different options and analyzing their potential improvement, I was able to make informed decisions. Each chosen solution was backed by solid reasoning, demonstrating a deep understanding of the project's requirements and constraints.

In summary, my frequent interactions with my supervisor, proactive approach, and thorough evaluation of alternatives have been essential in driving the project's progress and ensuring successful outcomes.