A picture containing tool

Description automatically generatedHakobyan Howhannes

Batch Size and Obstacle Avoidance:

A PPO-Based Study of Autonomous Vehicles Using Unreal Engine Learning Agents

Supervisor: Boury Fries

Coach: Geeroms Kasper

Graduation Work 2024-2025

Digital Arts and Entertainment

Howest.be

A close up of a card

Description automatically generated

Contents

[2. Abstract & Key words 3](#_Toc186482203)

[3. Preface 4](#_Toc186482204)

[4. List of Figures 5](#_Toc186482205)

[5. Introduction 6](#_Toc186482206)

[6. Literature Study / Theoretical Framework 7](#_Toc186482207)

[6.1 Introduction 7](#_Toc186482208)

[6.2 Machine Learning 7](#_Toc186482209)

[6.3 Reinforcement learning (RL) 8](#_Toc186482210)

[6.4 Proximal Policy Optimization 10](#_Toc186482211)

[6.5 Policy batch size 10](#_Toc186482212)

[6.6 Autonomous driving 11](#_Toc186482213)

[6.7 Unreal Engine Learning Agents 11](#_Toc186482214)

[7. Research 12](#_Toc186482215)

[7.1 Introduction 12](#_Toc186482216)

[7.2 Experimental Setup 12](#_Toc186482217)

[7.2.1 Framework 12](#_Toc186482218)

[7.2.2 Environment 12](#_Toc186482219)

[7.2.3 Agent 12](#_Toc186482220)

[7.3 Training 12](#_Toc186482221)

[7.3.1 Learning To Drive 12](#_Toc186482222)

[7.3.2 Obstacle Avoidance 12](#_Toc186482223)

[7.4 Testing 12](#_Toc186482224)

[7.4.1 Environment 12](#_Toc186482225)

[7.4.2 Data Collection 12](#_Toc186482226)

[7.5 Results 13](#_Toc186482227)

[8. Discussion 14](#_Toc186482228)

[9. Conclusion 15](#_Toc186482229)

[10. Future work 16](#_Toc186482230)

[11. Critical Reflection 17](#_Toc186482231)

[12. References 18](#_Toc186482232)

[13. Acknowledgements 19](#_Toc186482233)

[14. Appendices 20](#_Toc186482234)

# Abstract & Key words

**An abstract explains the outline of the paper concisely (the methods, results, etc.). Maximum length of 250 words, preferably both in English and Dutch.**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum ac quam nec arcu semper dignissim. Nulla quam magna, varius sit amet pharetra et, dictum quis elit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin ullamcorper, ipsum sit amet scelerisque rhoncus, leo quam rhoncus elit, sit amet ullamcorper tellus nisi eget sapien. Suspendisse potenti. Ut non justo viverra, tempus felis vitae, elementum mi. Morbi at dui sed lacus fringilla condimentum. Duis non odio ac arcu volutpat vehicula eu et turpis. Praesent vitae magna ante. Nulla in orci lacus. Donec quis vestibulum mi. Sed ipsum sapien, pretium maximus purus sed, bibendum consequat lectus. Aliquam porttitor dolor eu gravida vulputate. Vestibulum ut urna eget massa tincidunt ultricies. Morbi hendrerit sapien at diam tincidunt semper. Aliquam ut quam dictum quam maximus tempor sed at felis.

# Preface

***A preface is a statement of the author's reasons for undertaking the work and may include personal comments that are not directly relevant to other sections of the thesis or dissertation.* No word count limit.**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum ac quam nec arcu semper dignissim. Nulla quam magna, varius sit amet pharetra et, dictum quis elit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin ullamcorper, ipsum sit amet scelerisque rhoncus, leo quam rhoncus elit, sit amet ullamcorper tellus nisi eget sapien. Suspendisse potenti. Ut non justo viverra, tempus felis vitae, elementum mi. Morbi at dui sed lacus fringilla condimentum. Duis non odio ac arcu volutpat vehicula eu et turpis. Praesent vitae magna ante. Nulla in orci lacus. Donec quis vestibulum mi. Sed ipsum sapien, pretium maximus purus sed, bibendum consequat lectus. Aliquam porttitor dolor eu gravida vulputate. Vestibulum ut urna eget massa tincidunt ultricies. Morbi hendrerit sapien at diam tincidunt semper. Aliquam ut quam dictum quam maximus tempor sed at felis.

# List of Figures

**The list of figures lists the figures in the order in which they appear throughout the thesis. They may be numbered sequentially, or be subdivided following the chapters in which they appear.**

[*Figure 1: Different Types of ML (Experfy, n.d.)* 6](#_Toc186555066)

[Figure 2: Reinforcement Learning framework (Scribbr, n.d.) 8](#_Toc186555067)

[Figure 3: Reinforcement Learning taxonomy (OpenAI, 2018) 9](#_Toc186555068)

[Figure 4: Track layout used for training AVs 11](#_Toc186555069)

# Introduction

**In the introduction, you write the background of your topic and discuss the observation that spurred you on to do this research project. Explain the purpose of the paper and present your research question(s) and the hypothesis at the end of this section. This section is typically a couple of pages long.**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum ac quam nec arcu semper dignissim. Nulla quam magna, varius sit amet pharetra et, dictum quis elit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Proin ullamcorper, ipsum sit amet scelerisque rhoncus, leo quam rhoncus elit, sit amet ullamcorper tellus nisi eget sapien. Suspendisse potenti. Ut non justo viverra, tempus felis vitae, elementum mi. Morbi at dui sed lacus fringilla condimentum. Duis non odio ac arcu volutpat vehicula eu et turpis.

Praesent vitae magna ante. Nulla in orci lacus. Donec quis vestibulum mi. Sed ipsum sapien, pretium maximus purus sed, bibendum consequat lectus. Aliquam porttitor dolor eu gravida vulputate. Vestibulum ut urna eget massa tincidunt ultricies. Morbi hendrerit sapien at diam tincidunt semper. Aliquam ut quam dictum quam maximus tempor sed at felis.

# Literature Study / Theoretical Framework

## Introduction

This section provides an overview of existing research and theoretical concepts relevant to the impact of experience replay batch size and obstacle avoidance AVs trained using PPO. It establishes the background needed to contextualize this study.

## Machine Learning

Machine learning (ML) enables computers to perform actions without explicit programming (Bi et al., 2019). When complex behaviors are required, it is often a burdensome task to explicitly program those behaviors. ML provides a more efficient approach to developing advanced and human-like behaviors.(Janiesch et al., 2021)

There are several types of ML, such as Supervised ML, Unsupervised ML, Reinforcement Learning (RL) and Semi-Supervised ML.(Jain, 2023). This paper will only explore RL, as it is a common choice when training AVs (Kiran et al., 2021)

*Figure 1: Different Types of ML (Experfy, n.d.)*



## Reinforcement learning (RL)

Reinforcement Learning (RL) is an area of machine learning where the agent develops its behavior based on actions, observations and rewards. A common way of achieving this is by trial-and-error. The agent is forced to interact with the environment, and based on the actions it performs, a positive reward or a negative reward, also referred to as a penalty, will be given. The goal is to maximize the total reward, while balancing exploration (trying new actions) and exploitation (choosing the best-known and most-rewarding action).(Crespo & Wichert, 2020)

The environment in which the agent is being trained can be extensive, with numerous observable elements and a wide range of possible actions. It is the responsibility of the developer to specify what an agent observes, the actions it can take, and the rewards for specific behaviors.

Here are some key concepts in RL:

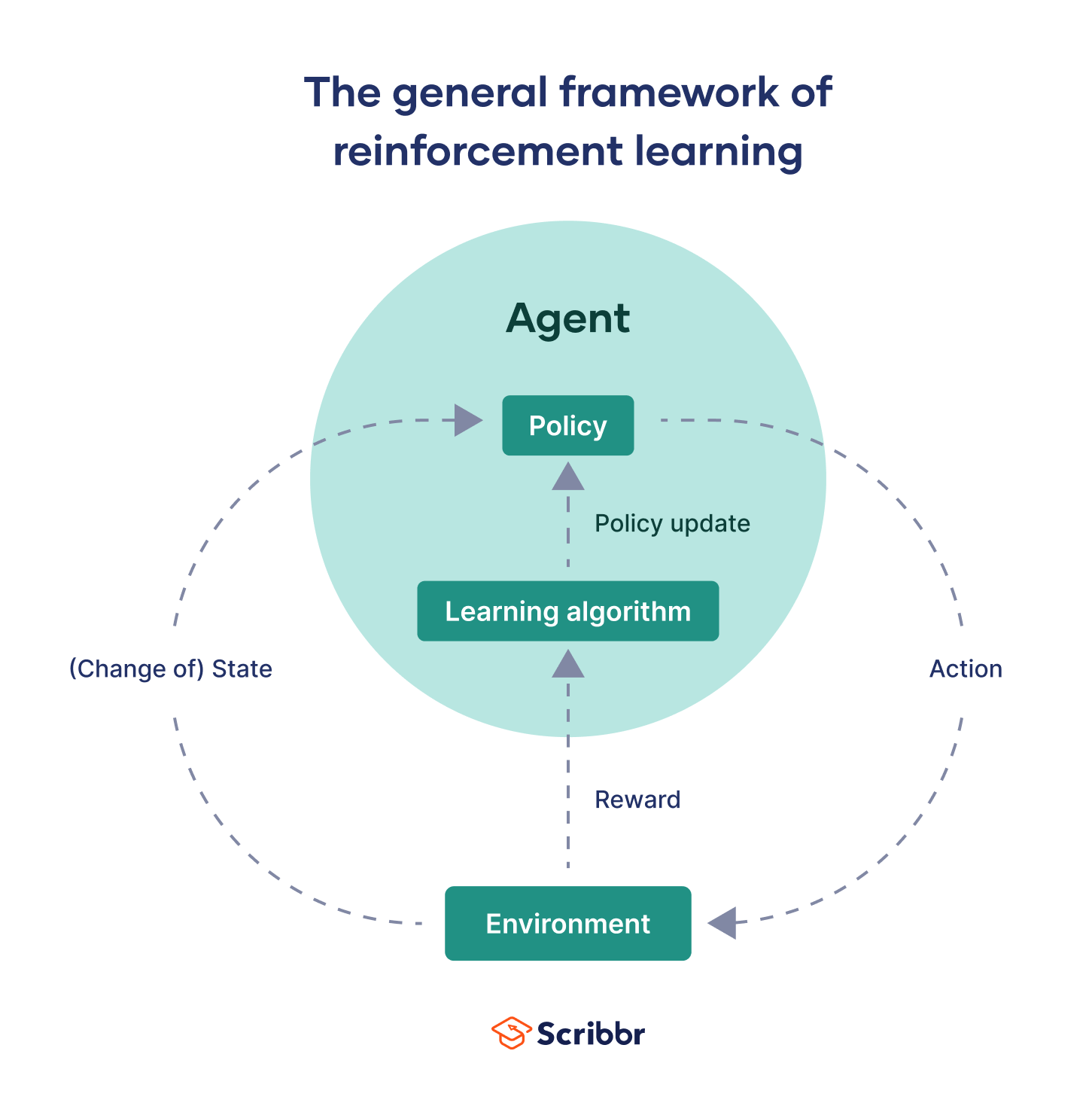
* *The State* represents the current condition of the environment. It encompasses all the information about the world. (OpenAI, 2018)
* *Observations* represent the information the agent receives from the environment. Using observations the current state of the environment is captured, and this information is fed into the agent’s knowledge base. For example, in a traffic situation, observations might include the positions, directions and speeds of nearby vehicles, the weather conditions, the status of traffic lights and the agent’s own velocity and direction. These observations heavily influence the agent’s decision-making (Arabaci, 2023).
* *Actions* represent the interactions with the environment. The collection of all valid actions in an environment is called the **action space** (OpenAI, 2018). One of the main challenges is finding the perfect balance between actions that yield high results, and actions that have not been executed yet as part of the discovery process (Kiran et al., 2021). Revisiting the earlier traffic example, possible actions could include throttle, braking, and steering.
* *Policies*are the link between the current state of the environment the agent finds itself in, and the action it should take in that situation. It can be a function, a matrix, or an array of probabilities of each action being chosen (Crespo & Wichert, 2020). The words “agent” and “policy” are often used interchangeably because the policy is essentially the agent’s brain (OpenAI, 2018).
* *Rewards* play a significant role in the agent’s development. In each state, the agent receives a reward from the environment based on the action it took. If that action leads to desired outcomes, the reward is positive, otherwise the reward is negative. The agent strives to maximize the cumulative rewards received over the span of its training, leading to desired behavior (Kiran et al., 2021). In the previously mentioned traffic scenario, actions such as driving through a red light, or crashing into other cars could yield negative rewards, while driving under the speed limit, and avoiding collisions could lead to positive rewards.

As previously mentioned, reward shaping is crucial in RL. A well-designed reward function will lead to a more desirable outcome, while a bad design might lead to suboptimal behavior. To shape the reward function properly, one must define the goal of the agent, and the key behaviors the agent must learn. Afterwards, all the actions or states that either reward or penalize the agent must be identified.

* **Rewards** are given to encourage the agent to take certain actions in specific scenarios
* **Penalties** are given to discourage the agent from taking incorrect actions.

It is also important to note that having a disproportionately high or low rewards or penalties can lead the agent to over-focus on certain actions, not prioritizing reaching the goal (GeeksforGeeks, 2024).

Figure : Reinforcement Learning framework (Scribbr, n.d.)

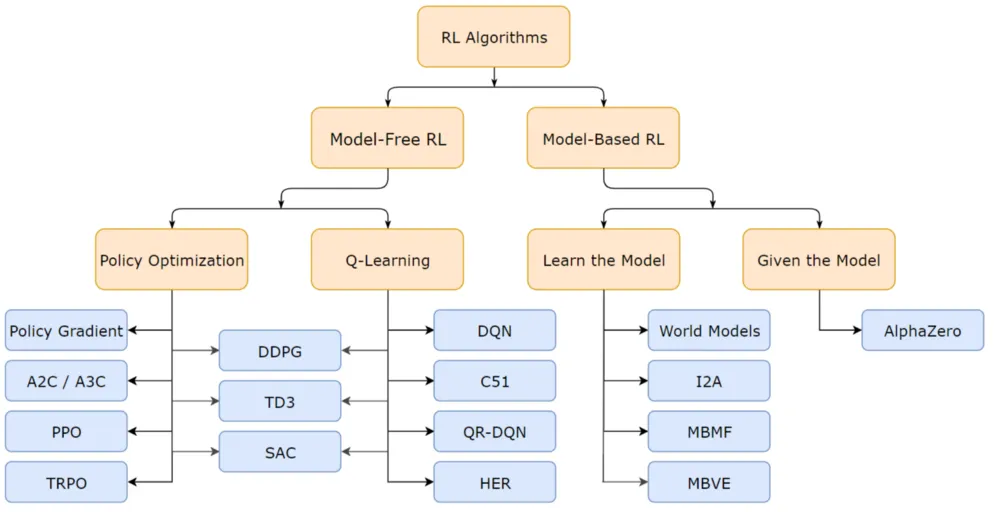


## Proximal Policy Optimization

Proximal Policy Optimization (PPO) is one of many RL algorithms, developed by John Schulman in 2017. It is widely used by American artificial intelligence company OpenAI and is considered a state-of-the-art RL algorithm by many experts (Kumar, 2024; OpenAI, 2017).

PPO is part of Policy Optimization methods, where the agent learns the **policy function** that maps state to action directly (SmartLab, 2019). To determine how well an agent is performing, and to decide which action the agent should take for each state, the policy function must be evaluated. When the agent is in the initial stages of learning and doesn’t know which actions lead to the best results, it does so by calculating the **policy gradients.** The output of the gradients helps the agent determine which actions are most likely to lead to higher rewards (Kumar, 2024).

Figure : Reinforcement Learning taxonomy (OpenAI, 2018)



## Policy batch size

**Policy batch size** refers to the number of samples processed together during training. In the context of RL, it specifically denotes the number of experiences grouped for training updates. Each experience typically comprises:

* The **state** of the agent
* The **action** the agent takes in that state
* The **reward** of the set action
* The **next state** of the agent
* Weather the episode terminates (Bacancy, 2024).

Before calculating a gradient update, a batch is processed. The size of this batch is controlled by the developer. For performance reasons it is advisable to keep the batch size to a power of two (jcm69, 2017).

Batch size is an important hyperparameter to tune. It is known that too large of a batch size will lead to poor generalization, and a smaller batch size is not guaranteed to deliver the desired outcome. It is also known that a smaller batch size leads to faster training, while a relatively big batch size leads to “guaranteed” results. It is important to mention that making general statements about the effects of hyperparameters is very difficult, as behaviour often varies from dataset to dataset (Shen, 2018). Therefore, it is up to the developer to determine what is referred to as the “optimal batch size” for the training process.

## Autonomous driving

Autonomous driving (AD) refers to the process of navigating a vehicle with partial or no human intervention. AVs utilize technology to avoid road hazards and adapt to traffic conditions (CFSS, n.d.). It has been a popular research topic since the end of the last century, because it promises many benefits such as safety and time savings (Wen et al., 2020).

Thanks to the rapid development of hardware, simulations enable us to test various AD algorithms with significantly reduced labour cost and time. More importantly, it takes the human error of breaking expensive equipment out of the equation (Wen et al., 2020).

Decision-making systems for real-life AV are often complex, including various sensors to fetch data from the environment, mapping algorithms for localization, understanding human drivers and predicting their behaviour, and engaging in kinematic manoeuvres (Barla, 2022). However, for the purposes of this paper, it is sufficient to focus solely on environment sensing.

The most common sensors used in real-life AV trainings include cameras, LiDAR, and RADAR (Barla, 2022). However, for the scope of this paper, ray tracing will be utilized for environment sensing and obstacle detection, as it is a more practical choice for this study.

## Unreal Engine Learning Agents

Learning Agents is an Unreal Engine ML plugin for AI bots. It currently supports reinforcement & imitation learning to create game-playing agents, physics-based animations, automated QA bots and more. It is not a general-purpose ML framework, so generative and conversational AI are not supported (Epic, 2024).

Learning Agents is intended for developers with different ML knowledge levels. Beginners can take advantage of the available plugins in Blueprints, while more advanced users can utilize the C++ API for greater control and flexibility (Epic, 2024; HuggingFace, n.d.).

The plugin comes with a built-in PyTorch PPO algorithm for RL. Functionality for defining observations, actions and rewards are provided (Epic, 2024).

To enhance the readability of this paper, the Blueprint interface will be used to implement functionalities and demonstrate key concepts.

# Research

## Introduction

The purpose of this research is to train autonomous vehicles (AVs) capable of navigating a road, staying within bounds, and avoiding static obstacles randomly placed along the path. Five AVs will be trained using identical hyperparameters, with the only variable being the policy batch size. After training, the AVs will be tested on three distinct track types: linear, circular, and zigzag. During these test episodes, data on obstacle avoidance performance will be collected and analyzed. The findings will be compared to draw meaningful conclusions. The detailed methodology is explained in the following sections.

## Experimental Setup

All the experiments in this paper are conducted in Unreal Engine (UE) 5.4. The engine offers a built-in vehicle template and a racing track map, which significantly reduces development time. Another key reason for choosing UE is to explore and test the Learning Agents (LA) plugin, as it is relatively new and underrepresented in academic research.

The Landscape tool in UE enables the creation of diverse track layouts with ease, saving significant time when designing multiple track configurations. Additionally, UE provides powerful debugging tools that are particularly beneficial for RL, where debugging can often be a challenging process.

This project utilizes LA 5.4. Despite being an experimental plugin, it perfectly serves the purposes of this research.

The default track layout, as shown in Figure 4, is used to carry out the training iterations. This layout is chosen as it represents a balanced middle ground, allowing for evaluation of the agent's behavior in both simpler and more complex maps after training. The length of the track is 1017m.

Figure : Track layout used for training AVs

A video game screen shot

Description automatically generated

UE SportsCar\_Pawn, that comes with the vehicle template, is utilized for both training and testing. No changes have been made to the vehicle set-up.

Figure : Vehicle used for training and testing



Five agents are trained over 5,000 iterations, each with a distinct batch size: 16, 32, 64, 128, and 256. To accelerate the training process, multiple instances of the same agent are trained simultaneously. The training is conducted in five phases, with each phase focusing on one batch size and involving 16 agents training concurrently. Upon completion of each phase, the training is halted, and the training data is saved into four Unreal Data Assets: Critic, Decoder, Encoder and Policy. TO BE CONTINUED

## Training

The training of a single agent is divided into two stages. In the first stage, the agent's sole focus is to complete a full lap while staying within the track boundaries. Once this is successfully achieved and the implementation is verified, static obstacles are introduced into the environment. In the second stage, the agent's behavior is updated to account for these obstacles, and the new goal becomes navigating the track while avoiding collisions with the obstacles.

### Learning To Drive

### Obstacle Avoidance

## Testing

Mention testing details here

### Environment

### Data Collection

## Results

# Discussion

**In this section, you offer an interpretation of the results you obtained and try to relate them to the theoretical framework you presented. This is typically not a very long section, but obviously one of the most important ones.**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum ac quam nec arcu semper dignissim. Nulla quam magna, varius sit amet pharetra et, dictum quis elit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Proin ullamcorper, ipsum sit amet scelerisque rhoncus, leo quam rhoncus elit, sit amet ullamcorper tellus nisi eget sapien. Suspendisse potenti. Ut non justo viverra, tempus felis vitae, elementum mi. Morbi at dui sed lacus fringilla condimentum. Duis non odio ac arcu volutpat vehicula eu et turpis.

Praesent vitae magna ante. Nulla in orci lacus. Donec quis vestibulum mi. Sed ipsum sapien, pretium maximus purus sed, bibendum consequat lectus. Aliquam porttitor dolor eu gravida vulputate. Vestibulum ut urna eget massa tincidunt ultricies. Morbi hendrerit sapien at diam tincidunt semper. Aliquam ut quam dictum quam maximus tempor sed at felis.

# Conclusion

**In this section, you ascertain the demonstrable outcomes of your study and outline the merits of the project for the academic field and the discourse community. This is typically not a very long section, but obviously also one of the more important ones.**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum ac quam nec arcu semper dignissim. Nulla quam magna, varius sit amet pharetra et, dictum quis elit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Proin ullamcorper, ipsum sit amet scelerisque rhoncus, leo quam rhoncus elit, sit amet ullamcorper tellus nisi eget sapien. Suspendisse potenti. Ut non justo viverra, tempus felis vitae, elementum mi. Morbi at dui sed lacus fringilla condimentum. Duis non odio ac arcu volutpat vehicula eu et turpis.

Praesent vitae magna ante. Nulla in orci lacus. Donec quis vestibulum mi. Sed ipsum sapien, pretium maximus purus sed, bibendum consequat lectus. Aliquam porttitor dolor eu gravida vulputate. Vestibulum ut urna eget massa tincidunt ultricies. Morbi hendrerit sapien at diam tincidunt semper. Aliquam ut quam dictum quam maximus tempor sed at felis.

# Future work

This paper explored one approach to developing autonomous driving using Reinforcement Learning. In machine learning, parameter tuning plays a critical role, so future work could focus on optimizing the balance between rewards and penalties to further enhance obstacle avoidance. Another potential direction could involve integrating dynamic obstacle avoidance and analysing how batch size influences performance in such scenarios.

Additionally, expanding the scope of observations and actions, by for example incorporating traffic conditions, could provide valuable insights into more complex driving behaviours. As the Learning Agents plugin continues to evolve, new features and possibilities are likely to emerge, opening further avenues for exploration and innovation in autonomous driving research.

# Critical Reflection

I began this research with no prior knowledge of Machine Learning. Through extensive reading and practical experimentation, I have developed a strong understanding of Reinforcement Learning, particularly how it can be applied to train agents for use in games. While my experiments focused specifically on autonomous driving, I’ve realized that the principles of Reinforcement Learning, namely defining observations, actions, and rewards, are broadly applicable across different domains.

This research has inspired me to explore Learning Agents further, experimenting with their capabilities beyond the scope of this project. Given that the Learning Agents plugin is still in an experimental stage, I plan to closely follow its development and adapt to its evolving features.

Looking ahead, I am excited about the potential for Reinforcement Learning to become more integrated into everyday applications. This research has been both challenging and rewarding, providing me with valuable insights and a strong foundation to build upon in future projects.

# References

**In this section, you list all the references you made in alphabetical order; consequently adhere to the referencing style you have chosen.**

Casey Raes (2014), Processing (second edition).

Saccade. (n.d.). In Wikpedia. Retrieved November 6 2016 from <https://en.wikipedia.org/wiki/Saccade>

Sarah Northway (2016) A year in VR Northway [Powerpoint slides] from <https://www.gdcvault.com/play/1024631/A-Year-in-VR-A>

# Acknowledgements

I would like to thank Alexandros Kougentakos for his invaluable assistance in troubleshooting my obstacle avoidance system. I am also deeply grateful to Kasper Geeroms, my coach, for his support and to Fries Boury, my supervisor, for his continuous guidance throughout this project.

# Appendices

**In many cases, there are items that were developed for a research paper that can’t go into the actual paper in full. Things suc as code, art pieces, output of statistical analysis, questionnaires, … In this section, you can present these elements; use the first page to list and number the items, then paste them sequentially. If some items are too large, you can store them online, and link to them. Common practice is to keep those links active at least one year after the publication of the thesis.**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum ac quam nec arcu semper dignissim. Nulla quam magna, varius sit amet pharetra et, dictum quis elit. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin ullamcorper, ipsum sit amet scelerisque rhoncus, leo quam rhoncus elit, sit amet ullamcorper tellus nisi eget sapien. Suspendisse potenti. Ut non justo viverra, tempus felis vitae, elementum mi. Morbi at dui sed lacus fringilla condimentum. Duis non odio ac arcu volutpat vehicula eu et turpis. Praesent vitae magna ante. Nulla in orci lacus. Donec quis vestibulum mi. Sed ipsum sapien, pretium maximus purus sed, bibendum consequat lectus. Aliquam porttitor dolor eu gravida vulputate. Vestibulum ut urna eget massa tincidunt ultricies. Morbi hendrerit sapien at diam tincidunt semper. Aliquam ut quam dictum quam maximus tempor sed at felis.